Clever Girl: Benevolent Sexism and Cardiovascular Threat

Veronica M. Lamarche

University of Essex

Mark D. Seery, Cheryl L. Kondrak, Thomas L. Saltsman, & Lindsey Streamer

University at Buffalo, The State University of New York

Correspondence concerning this article should be addressed to Veronica M. Lamarche, Department of Psychology, Wivenhoe Park, University of Essex, Colchester, Essex, United Kingdom, CO4 3SQ. E-mail: v.lamarche@essex.ac.uk

ACCEPTED VERSION

Published version available here:
https://www.sciencedirect.com/science/article/pii/S030105111830807X?via%3Dihub

Lamarche, V. M., Seery, M. D., Kondrak, C. L., Saltsman, T. L., & Streamer, L. (2019). Clever Girl: Benevolent Sexism and Cardiovascular Threat. *Biological Psychology*, 107781.

Abstract

Benevolent sexism is a double-edged sword that uses praise to maintain gender inequality, which consequently makes women feel less efficacious, agentic and competent. This study investigated whether benevolently sexist feedback that was supportive could result in cardiovascular responses indicative of threat (lower cardiac output/higher total peripheral resistance). Women received either supportive non-sexist or supportive yet benevolent sexist feedback from a male evaluator following practice trials on a verbal reasoning test. As expected, women receiving benevolent sexist feedback exhibited cardiovascular threat during a subsequent test, relative to women receiving non-sexist feedback. There was no support for an alternative hypothesis that benevolent sexist feedback would lead to cardiovascular responses consistent with disengaging from the task altogether (i.e., lower heart rate and ventricular contractility). These findings illustrate that the consequences of benevolent sexism can occur spontaneously, while women are engaged with a task, and when the sexist feedback is intended as supportive.

Keywords: benevolent sexism, challenge/threat, cardiovascular reactivity

Clever Girl: Benevolent Sexism and Cardiovascular Threat

Despite best efforts to decrease gender bias and inequality in the past decades, women worldwide continue to experience sexism in one way or another (Kehn & Ruthig, 2013; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012). Whether these experiences are explicitly derisive, such as implying a female colleague received a big promotion because of her appearance, or relatively more benign, assuming a female colleague would not be interested in joining the office Fantasy Football league thus inadvertently excluding her from an informal networking opportunity, they have tangible consequences for women in all walks of life. In fact, even covert manifestations of sexism have been linked to serious social issues including barriers that make it more difficult for women to enter or remain in the workforce, victim blaming following sexual assault, and stifled social change (Abrams, Viki, Masser, & Bohner, 2003; Becker & Wright, 2011; Glick & Fiske, 2001; King et al., 2012; Viki & Abrams, 2002). Thus, the consequences of covert forms of sexism remain serious even though this type of sexism is often dismissed as well-intentioned, likeable, and even socially acceptable and desirable (Barreto & Ellemers, 2005; Bohner, Ahlborn, & Steiner, 2010; Moya, Glick, Exposito, de Lemus, & Hart, 2007). However, the majority of this research has focused on self-reported consequences of covert sexism, which rely on introspection in order to articulate the personal impact of the experience. The aim for this study was to test whether being on the receiving end of covert forms of sexism that are often disguised as helpful and supportive can have immediately negative consequences for women engaged in an activity related to the sexist comments, even when they do not have the opportunity for conscious rumination and deliberation. In order to address this limitation, we relied on cardiovascular indicators of challenge/threat psychological states while women worked on a verbal reasoning task after receiving benevolently sexist feedback.

Malevolent Benevolence

Sexism is represented by two complementary ideological support systems (Glick & Fiske, 1996; Glick et al., 2000). People, both men and women, who endorse hostile sexist beliefs hold negative and antagonist attitudes towards women, labelling them as categorically inferior to men. This overt form of sexism is easily identified and called out (Barreto & Ellemers, 2005; Becker & Wright, 2011). If hostile sexism represents overt, derisive and antagonistic attitudes towards women, benevolent sexism is its kinder, gentler cousin. As a result, experiencing benevolent sexism can be disorienting, as it may appear supportive while simultaneously reinforcing and extoling the virtues of traditional gender roles, responsibilities, and capabilities. For example, client portfolio managers Karen and Ben have been eagerly awaiting their annual performance reviews from their manager. While Karen's review praises her for being a likeable team player that people enjoy working with, Ben's focuses on his efficacy as a team leader and tough negotiation skills. Although both of these reviews are positive, Karen is left without any constructive feedback on her skills that can help her develop her career and neglects to reinforce her competency in her current role. Benevolent sexism (BS) maintains gender inequality through the idealization and assertion of subordinate qualities in women (Glick & Fiske, 1996). By focusing on qualities that, while potentially positive, hold less social power and capital, BS affords men the means of offering support to women, while still maintaining the status quo of traditional gender hierarchies (Becker & Wright, 2011; Glick & Fiske, 1997, 2001; Glick et al., 2000; Jost & Kay, 2005; Lee, Fiske, Glick, & Chen, 2010). The relatively positive tone of BS means that perpetrators are often seen in a positive light and are even less likely to be labeled as sexist (Barreto & Ellemers, 2005; Goh & Hall, 2015). Conversely, women who reject BS support are judged as cold and uncaring (Becker, Glick, Ilic, & Bohner, 2011). Thus, the more palatable

nature of BS means that it often goes unnoticed, unchecked and unchallenged.

But just like hostile sexism, BS negatively impacts women's success and well-being. These negative outcomes are not limited to situations in which women are perceived unfavorably for rejecting BS support. Rather, BS itself confines women to a pedestal built on the idyllic notion of traditional gender roles, with men serving as chivalrous protectors and providers (Viki & Abrams, 2003; Overall, Sibley & Tan, 2011). This gallantry comes at the cost of ignoring women's abilities, instead offering ineffectual and patronizing support (Hammond & Overall, 2015; Shnabel, Bar-Anan, Kende, Bareket & Lazar, 2016). As a result, women who experience BS show decreased efficacy, performance, agency and competency (Jones et al., 2014; Dardenne, Dumont, & Bollier, 2007; Dumont, Sarlet, & Dardenne, 2010). For example, women got fewer questions right on a problem-solving test framed as a hiring tool when the ostensible recruiter expressed benevolent sexist attitudes towards women, and this impaired performance was further linked to mental intrusions these women had about their competence (Dardenne et al., 2007). These negative behaviors and cognitions triggered by BS are attributed to learned helplessness, as women feel that they do not have the resources to combat the patronizing nature of the support that simultaneously praises and devalues them (Vescio, Gervais, Snyder & Hoover, 2005).

To date, research suggests that the negative consequences of BS are due to rumination on and suppression of thoughts related to incompetency (Dardenne et al., 2007, 2013). Rumination has been associated with impaired cardiovascular recovery from negative emotional stressors (Key, Campbell, Bacon, & Gerin, 2008; Glynn, Christenfeld, & Gerin, 2002). Consistently, benevolent sexism, but not hostile sexism, was linked to delayed cardiovascular recovery in a study by Salomon, Burgess, and Bosson (2015), in which women were told that the difficult part

of a performance task would be removed because women struggle completing the difficult question (hostile sexism condition), women dislike and are upset by the hard section (benevolent sexism condition), or that it was randomly determined by the computer (control). Women in the benevolent sexism condition exhibited poorer cardiovascular recovery (i.e., return to baseline) following the task compared to those in the hostile sexism and no-sexism control conditions. Likewise, women in both sexism conditions scored higher on ruminative thoughts compared to the controls, although there was no mediating effect of rumination on recovery. Differences in cardiovascular reactivity during the task between the benevolent sexism and control conditions (most closely paralleling the current focus) were, excepting one measure, absent. However, one limitation of this study was that the sexist feedback the women received in both the hostile and benevolent sexism conditions was not based on their own performance or abilities, but instead assumptions about women in general.

Thus, what still remains unresolved across the body of research examining responses to BS feedback is whether the negative impact of BS on feelings of competence can occur (1) spontaneously, without undistracted opportunity for rumination or deliberation, (2) *during* subsequent active task performance, and (3) specifically, after BS performance feedback that is encouraging of women's ability to perform well in the future (vs. comments that discount ability in anticipation of performance; Salomon et al., 2015). To address these issues, we used cardiovascular measures sensitive to evaluations of personal resources and situational demands, using the biopsychosocial model of challenge/threat as a theoretical framework.

The Biopsychosocial Model of Challenge/Threat

Physiological indicators of psychological responses make it possible to tap into a person's psychological state while they are actively engaged in a task, thus eliminating the

disadvantages of interrupting participants to assess how they feel during the task, or relying on potentially inaccurate reflection upon completion. The biopsychosocial model of challenge/threat (BPSC/T; Blascovich, 2008; Blascovich, & Tomaka, 1996; Seery, 2011, 2013; Seery, & Quinton, 2016) allows for insight into people's evaluations of personal coping resources and situational demands by assessing cardiovascular responses during motivated performance situations, in which they are actively working to pursue a self-relevant goal (e.g., completing an intelligence test that will be evaluated). The evaluations of resources and demands do not necessarily happen through deliberative or conscious processes. For instance, stimuli presented outside of conscious awareness have been shown to affect relevant cardiovascular responses (Weisbuch-Remington, Mendes, Seery, & Blascovich, 2005), and these cardiovascular responses have correlated with relatively uncontrollable nonverbal behaviors rather than controllable ones (Weisbuch, Seery, Ambady, & Blascovich, 2009). Furthermore, the evaluation process is thought to be dynamic and occur fluidly, influenced by a range of interrelated factors such as the presence of others and familiarity (for discussion see Blascovich, 2008; Seery, 2011). Relevant for the current study, task instructions and verbal feedback have been used to manipulate evaluations and the accompanying cardiovascular responses (for examples see Moore, Vine, Wilson, & Freeman, 2012; Tomaka, Blascovich, Kibler, & Ernst, 1997; Turner, Jones, Sheffield, Barker, & Coffee, 2014). Thus, these processes make the model ideal for assessing the consequences of benevolent sexist feedback on personal resource versus situational demand evaluations during a subsequent, ongoing motivated performance situation.

According to the BPSC/T model, *task engagement* reflects the extent to which the goal is considered personally relevant or subjectively valuable. The psychological states of challenge and threat are determined by the evaluation of personal coping resources relative to situational

demands in that moment. Relative *challenge* occurs when a person evaluates high personal resources and low situational demands (consistent with self-confidence, Weisbuch et al., 2009; Williams, Cumming, & Balanos, 2010). Conversely, relative *threat* occurs when evaluated resources are low and demands are high (consistent with self-doubt, Seery, Blascovich, Weisbuch, & Vick, 2004). Despite the discrete labels of challenge and threat, they in fact represent anchors on a single bipolar continuum, such that greater relative challenge reflects feeling more capable of handling the situational demands, whereas relative threat reflects feeling less capable. Furthermore, consistent with psychological assessments of self-confidence and capability, cardiovascular challenge is associated with better task performance relative to cardiovascular threat states (Behnke & Kaczmarek, 2018; Hase et al., in press). Thus, when a woman finds herself in a situation that leads her to experience self-doubt, such as feeling less competent when attempting to complete an evaluated task, she should experience relative threat and exhibit the associated cardiovascular pattern.

There are four cardiovascular indexes of challenge/threat that are assessed during a motivated performance situation: (1) heart rate (HR); (2) preejection period (referred to here as ventricular contractility; VC), a measure of the left ventricle's contractile force (for presentational purposes, VC = reactivity × −1); (3) cardiac output (CO), the amount of blood pumped by the heart per minute; and, (4) total peripheral resistance (TPR), a measure of net constriction versus dilation in the arterial system. The use of these measures builds on work from Dienstbier (1989; for additional discussion, see Seery, 2011), which argues that the body mobilizes energy reserves to respond to motivated performance situations via the activation of the sympathetic-adrenomedullary (SAM) and pituitary-adrenocortical (HPA or PAC) axes. SAM axis activation leads to the release of catecholamines, epinephrine and norepinephrine, which

because of their short half-lives make a short spike of activation possible, a response that is associated with positive outcomes in the face of stressors (Dienstbier, 1989). Conversely, HPA axis activation leads to the slow release of longer-lasting cortisol, a pattern that is associated with relatively negative outcomes. Task engagement during a motivated performance situation is believed to result in increased SAM activation, which leads to increases in HR and VC from baseline. Larger increases in HR and VC reflect greater task engagement (e.g., Blascovich, Mendes, Hunter, & Salomon, 1999; Seery, Weisbuch, & Blascovich, 2009; also see Fowles, Fisher, & Tranel, 1982; Tranel, Fisher, & Fowles, 1982; for additional discussion, see Seery, 2013). Because task engagement is a necessary precursor for challenge/threat, both challenge and threat responses are marked by increases in HR and VC.

Challenge and threat are therefore differentiated by differences in CO and TPR: challenge leads to higher CO and lower TPR than threat. Once engaged in a task, challenge is believed to result in the preferential release of epinephrine from the adrenal medulla. Via action on beta-2 receptors, epinephrine results in dilation of arteries in large skeletal muscles (lower TPR); coupled with the HR and VC increases common across the challenge/threat continuum, this facilitates the heart in pumping more blood (higher CO; Brownley, Hurwitz, & Schneiderman, 2000; Papillo & Shapiro, 1990). In contrast, threat is thought to lead to elements of both SAM and HPA activation, diverging from challenge by inhibiting epinephrine release in particular and thus its vasodilatory effects (see Seery, 2011, 2013). The validity of these cardiovascular markers has been supported by dozens of individual studies which have assessed or manipulated challenge/threat states in different ways (e.g., Moore et al., 2012; Moore, Vine, Wilson, & Freeman, 2014; Moore, Wilson, Vine, Coussens, & Freeman, 2013; Scheepers, de Wit, Ellemers, & Sassenberg, 2012; Shimizu, Seery, Weisbuch, & Lupien, 2011; Tomaka, Blascovich, Kelsey,

& Leitten, 1993; Tomaka et al., 1997; Turner et al., 2014; Weisbuch-Remington et al., 2005), as well as in several recent reviews assessing the robustness and reliability of these effects (for examples see: Behnke & Kaczmarek, 2018; Blascovich, 2008; Hase et al., 2018; Seery, 2013). Although the model's cardiovascular measures are still susceptible to error like all measures (Behnke & Kaczmarek, 2018), and do not equate to the model's psychological states but instead imperfectly reflect them, the cardiovascular indexes of challenge/threat should still provide the means of assessing whether the negative psychological consequences of BS are spontaneous and occur immediately during performance.

Overview & Hypotheses

Previous research has shown that BS undermines women's feelings of competency. We used cardiovascular indexes of challenge/threat to assess women's responses during performance of a verbal reasoning task, after first receiving encouraging feedback in the performance domain that contained either BS or non-sexist content. By using cardiovascular measures that enabled us to differentiate between task engagement, challenge, and threat responses, the BPSC/T model allowed us to test two possible outcomes of supportive BS feedback. First, if women internalize the task as important (e.g., their intelligence is being judged) but BS feedback leaves them feeling like their personal resources are unable to meet the situational demands (e.g., lower competence), they should experience high task engagement and threat. Thus, we hypothesized that compared to women who received non-sexist feedback, women who received BS feedback would exhibit cardiovascular responses consistent with comparable task engagement (increases in HR and VC from baseline that do not differ by condition) and relative threat (lower CO/higher TPR). Second, an alternative possibility is that feedback undermining competence may motivate women to defensively distance from the difficult task after an initial poor performance (e.g., just

an unimportant lab experiment), which may consequently lessen the importance of the subsequent task. This would result in low task engagement. Thus, our alternative hypothesis was that compared to women who received non-sexist feedback, women who received BS feedback would exhibit cardiovascular responses consistent with lower task engagement (lower HR/VC). Either of these cardiovascular patterns would be consistent with the purported psychological consequences of BS feedback, including lower competency and learned helplessness. Although we believed a threat response should be more consistent with how women likely experience sexist feedback in performance situations in everyday life (i.e., they continue to value their work but the feedback they receive undermines their ability to succeed at it), we tested both possibilities. Finally, cardiovascular threat has been associated with worse performance than challenge (e.g., Blascovich, Seery, Mugridge, Norris, & Weisbuch, 2004; Moore, Vine, Wilson, & Freeman, 2012). Thus, although secondary to our cardiovascular predictions, we further explored whether BS would lead to worse performance than non-sexist feedback.

Method

Participants

Eighty-seven women who were native English speakers participated in exchange for course credit (81% 18-19yo; 60% Caucasian). The following participants were excluded: two due to computer errors, two for unusable physiological data, one for not following instructions, one for reporting having heard the manipulation prior to participating, and eight for guessing the hypothesis of the study (responses to sexist feedback), leaving 73 participants. This should have provided power=.80 to detect an effect size of $\eta^2_{partial}$=0.10, and power=.50 to detect an effect size of $\eta^2_{partial}$=0.05.

Procedure

Study sessions were run by female experimenters to avoid intergroup performance effects (Scheepers, 2009). Participants first completed demographic questions, which also served as an acclimation period to the lab setting prior to recording physiological responses. Next, they were attached to physiological sensors and rested during a 5-min resting baseline period (see below). Participants then completed 12 items of a difficult version of a verbal reasoning task (Remote Associates Task; RAT; McFarlin & Blascovich, 1984) under the guise that it measured intelligence and predicted important life outcomes. Each item required participants to generate the single word that linked three stimulus words together. Participants had 15s to generate an answer before the presenting computer advanced to the next item (3mins total). Supporting the difficulty of the test, participants averaged 0.71 items correct out of 12. A verbal reasoning task was chosen because it is a stereotype-consistent domain for women (Davies, Spencer, Quinn, & Gerhardstein, 2002; Seibt & Förster, 2004, Skaalvik & Skaalvik, 2004), thereby avoiding potentially complicating stereotype-threat effects in stereotype-inconsistent domains (e.g., math), which can lead to cardiovascular threat (Vick, Seery, Blascovich, Weisbuch, 2008). Participants stated their answers aloud for the experimenter to record, and instructions encouraged guessing to fit the cover story that even incorrect answers could be evaluated. Participants were next told that their responses were being scored by the lead researcher in charge of the lab, who would provide them with feedback to help them succeed on the next round of the RAT. After a 5-min rest period, participants were randomly assigned to either the BS feedback or no sexism feedback condition. Following the feedback, participants completed another 12-item version of the RAT, this time of moderate difficulty to avoid the task feeling impossible despite optimistic feedback (i.e., with another difficult version of the task) or surprisingly easy compared to the first task (i.e., with an easy version of the task), during which their cardiovascular responses were

measured. Next, they completed measures assessing gender identification, gender-role attitudes and questions about the performance task. Finally, they were debriefed and thanked for their participation. The self-report measures used in this study can be reviewed in full in Appendix A and the pre-recorded audio-instructions in Appendix B. This research complied with APA ethical standards in the treatment of our human participants. The protocol was approved by the university's Social and Behavioral Research Office IRB and all participants gave informed consent prior to participating in the study.

Baseline Period. Following the initial in-lab acclimation period in which participants completed demographic measures, participants were attached to the non-invasive physiological sensors followed by a 5-minute resting baseline period. Five-minute baseline periods are commonly used and provide a long enough relaxation period without risking that participants become bored and fidgety, or begin to perseverate on the upcoming tasks, which plausibly occurs during longer baselines. Participants received the following recorded instructions: *"Please try to move as little as possible for the rest of the experiment, although you may move when necessary. When possible, rest your hands on top of the lap tray. Please do not touch the keyboard unless you are instructed to do so. Before the study begins, we need to calibrate our physiological equipment. This occurs entirely in the experimenter's control room, so it will seem to you like nothing is happening. All we need you to do is sit quietly. This will take a few minutes. For the next few minutes, please sit quietly and relax until the experimenter tells you it is time to continue."* Experimenters monitored the incoming signals from a control room, and the participants via a live video feed.

Sexist Feedback. Feedback was pre-recorded by one male researcher to ensure consistent delivery across participants. The female experimenter running the session told participants their

total score on the first RAT and then told them they were being connected to the lead researcher via intercom. In both conditions, a man delivered the pre-recorded feedback in a pleasant and supportive vocal tone.

All participants heard:

> *OK, so it looks like you struggled with this test so far. But the first set of questions were just practice questions, so the next set that you'll answer will be what you are evaluated on.*

Followed by either:

1. BS Feedback: *You seem like a **very smart girl** because your answers showed a lot of creativity. I know it's hard not to get **emotional** during this type of test, but I'm sure you'll do well on the next set of questions as long as you **don't let your nerves get the best of you**.*

Or:

2. Non-Sexist Feedback: *You seem like a **very smart person** because your answers showed a lot of creativity. I know it's hard to **come up with answers** during this type of test, but I'm sure you'll do well on the next set of questions as long as you **continue to think outside of the box**.*

Cardiovascular Measures. Cardiovascular measures were recorded noninvasively, following accepted guidelines (Sherwood et al., 1990). We used the following equipment manufactured and/or distributed by Biopac Systems, Inc (Goleta, CA): NICO100C impedance cardiography (ICG) noninvasive cardiac output module, ECG100C electrocardiogram (ECG) amplifier, and NIBP100A/B noninvasive blood pressure module. ICG signals were detected with a tetrapolar aluminum/mylar tape electrode system, recording basal transthoracic impedance (Z0) and the first derivative of impedance change (dZ/dt), sampled at 1kHz. ECG signals were detected using a Standard Lead II electrode configuration (additional spot electrodes on the right arm and left leg, with ground provided by the ICG system), sampled at 1kHz. The blood pressure monitor was wrist-mounted, collecting continual readings—every 10-15 seconds—from the

radial artery of participants' nondominant arm. In combination, ICG and ECG recordings allowed computation of HR, VC (for presentational purposes, pre-ejection period reactivity × -1), and CO; the addition of blood pressure monitoring allowed computation of TPR (mean arterial pressure × 80 / CO; Sherwood et al., 1990). Recorded measurements of cardiovascular function were stored on a computer and analyzed off-line with Biopac Acqknowledge 3.9.2 for Macintosh software, using techniques from previously published challenge/threat research with the same equipment configuration (e.g., Le et al., 2019; Lupien, Seery, & Almonte, 2012; Seery, Leo, Lupien, Kondrak, & Almonte, 2013; Shimizu, Seery, Weisbuch, & Lupien, 2011; see Seery, Kondrak, Streamer, Saltsman, & Lamarche, 2016, for additional details), including ensemble averaging in 60 s intervals (Kelsey & Guethlein, 1990). The approach was comparable to techniques used in other challenge/threat work (e.g., de Wit, Scheepers, & Jehn, 2012; Moore, et al., 2012, 2014; Turner, Jones, Sheffield, & Cross, 2012; Turner et al., 2013). Scoring of cardiovascular data was performed blind to condition and other participant data.

Cardiovascular reactivity values (Llabre, Spitzer, Saab, Ironson, & Schneiderman, 1991) were calculated by subtracting the last minute of baseline from task values. The last minute of baseline was selected as it should represent the truest resting state after relaxing for several minutes. The decision to use the final minute of baseline was determined a priori, in keeping with our previous work (e.g., Le et al., 2019; Saltsman, Seery, Kondrak, Lamarche, & Streamer, 2019; Seery et al., 2016; Streamer, Seery, Kondrak, Lamarche, & Saltsman, 2017). As is standard in challenge/threat research in general and when using this task in particular (e.g., Seery, Blascovich, Weisbuch, & Vick, 2004; Lupien, Seery & Almonte, 2012), our a priori strategy was to use the mean of reactivity from the first two task minutes in analyses. This strikes a balance between capturing maximal reactivity—which often occurs at the beginning of a

task—and incorporating multiple reactivity data points to enhance reliability, which relying only on the first task minute fails to do. Cardiovascular reactivity values that exceeded 3.3 SD from the grand mean (p=.001 in a normal distribution) were identified as extreme values (Tabachnick & Fidell, 1996) and were winsorized by recoding them to be 1% higher than the next-highest non-extreme value. This preserved the rank order of values in the distribution while decreasing the influence of extreme scores. A total of 4 values were changed using this procedure (1 for HR, 1 for CO, 2 for TPR).

Because the same physiological activation thought to underlie the differences between challenge/threat cardiovascular patterns (i.e., release of epinephrine from the adrenal medulla) should affect both TPR and CO reactivity (r=-0.73), such that both indicate relative differences in challenge/threat, we combined the two into a single index to maximize their reliability, as is standard practice (e.g., Blascovich et al., 2004; de Wit et al., 2012; Moore et al., 2012). Reactivity scores for TPR and CO were first converted to z-scores, TPR's z-score was reverse-scored (i.e., multiplied by -1) because TPR and CO are expected to respond in opposite directions, and then the z-scores were summed. This is analogous to averaging over multiple items in a single questionnaire measure. It provides additional advantages of simplifying analyses with a single test of challenge/threat and capturing the relative combined pattern of CO and TPR within participants (e.g., differentiating between low CO/high TPR and low CO/moderate TPR). For tests of task engagement, HR and VC were similarly combined into a single index (r=.47) by summing their z-scores. Both indexes were standardized for ease of interpretation (M=0, SD=1). Higher scores on the challenge/threat index reflected greater relative challenge and lower scores reflected greater relative threat (zero was simply the sample mean, not their demarcation point), whereas higher scores on the task engagement index reflected

greater engagement in the task at hand. Separate analyses of the index components (reported below) revealed the same pattern of results.

Task Performance. Performance on the test RAT following sexist feedback was calculated by summing the total number of correctly generated responses to the 12 prompts (in reality, each item had a single correct answer). Participants received a score of 1 for each correct response, for a minimum of 0 correct responses and a maximum of 12 correct responses.

Gender-Role Attitudes and Identification Covariates. Because verbal performance is a stereotype-consistent domain for women (Davies, Spencer, Quinn, & Gerhardstein, 2002; Seibt & Förster, 2004, Skaalvik & Skaalvik, 2004), it was plausible that higher endorsement of traditional gender roles and gender identification could be associated with more positive resource/demand evaluations and hence greater challenge during a verbal task, perhaps especially in a context that makes gender salient. We thus decided a priori to include measures of attitudes and identification as covariates in analyses. Gender identification was measured using 4-items modified from the Collective Self-Esteem Scale (Luhtanen & Crocker, 1992; "In general, being a woman/man is an important part of my self-image."; Likert scale, 1=strongly disagree, 7=strongly agree) and traditional gender-role attitudes were measured using the 17-item Sex-Roles Attitudes Scale (Van Yperen & Buunk, 1991; "It is best to maintain more or less traditional gender roles in a relationship."; Likert scale, 1=strongly disagree, 7=strongly agree). The covariates were measured at the end of the study to help conceal the research question as well as prevent the activation of gender and sexism before the manipulation.

Results

Analytic Strategy

We used the following approach to test the primary and secondary hypotheses for this study. First, we addressed our primary hypothesis, our favored version of which was that benevolent sexist feedback would lead to cardiovascular reactivity consistent with a relative threat response, using linear regression to compare cardiovascular responses between conditions, while controlling for individual differences in gender identification and gender role attitudes. Specifically, after confirming evidence for the task-engagement prerequisite for challenge/threat in a preliminary step, we began by testing the alternative hypothesis that BS feedback may lead women to defensively distance from the task at hand (defensive distancing hypothesis), followed by testing our hypothesis that BS feedback undermines personal resources (undermined competence hypothesis), and an examination as to whether gender identification and gender-role attitudes can moderate the impact of BS feedback on cardiovascular responses.

Next we addressed our secondary hypotheses that BS feedback should lead to reduced performance on the second performance task using regression analyses to compare the between-condition differences in mean number of correct responses during the test RAT, as well as compare self-reported perceptions of participants' performance.

Primary Analyses: Task Engagement and Challenge/Threat

Task engagement.

Evidence for task engagement in the sample. See Table 1 for correlations and descriptive statistics. Because increases in HR and VC from baseline are common to the cardiovascular patterns across the challenge/threat continuum, we first tested whether HR and VC reactivity was significantly greater than zero during the post-feedback RAT with one-sample *t* tests. Results confirmed that for the sample as a whole, both HR and VC increased significantly

from baseline: HR M=4.13, t(72)=6.83, p<.001; VC M=3.54, t(72)=4.60, p<.001. This justified testing for differences in challenge/threat.

Defensive distancing hypothesis. Our alternative hypothesis was that BS feedback may lead to low task engagement if it motivates women to defensively distance from the difficult task after an initial poor performance. We thus next tested for differences in cardiovascular task engagement responses as a function of condition, controlling for gender-role attitudes and gender identification. This revealed no significant differences in the cardiovascular responses consistent with task engagement, including the task engagement index (b=0.11, t(69)=.47, p=.63), HR (b=1.26, t(69)=1.01, p=.32), and VC (b=-0.06, t(69)=-.03,p=.97). There was therefore no support for the alternative hypothesis that BS feedback leads women to defensively distance from the performance domain by rendering it unimportant, thus resulting in less task engagement compared to women who received non-sexist feedback.

Table 1
Correlations and Descriptive Statistics

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Challenge/threat index	--									
2. CO reactivity	.93***	--								
3. TPR reactivity	-.93***	-.73***	--							
4. Task engagement index	.21	.38***	-.01	--						
5. HR reactivity	-.06	.09	.21	.84***	--					
6. VC reactivity	.41***	.54***	-.22	.86***	.47***	--				
7. Gender identification	.18	.16	-.18	-.03	-.004	-.03	--			
8. Gender-role attitudes	.37**	.35**	-.33**	.13	.02	.19	-.004	--		
9. RAT #1 performance	-.16	-.11	.18	.03	.03	.02	.13	-.01	--	
10. RAT #2 performance	-.01	-.01	.002	-.17	-.22	-.04	-.13	-.03	.33**	--

M	0.00	-.47	100.63	0.00	4.13	3.39	5.06	3.34	.70	5.41
SD	1.00	1.30	145.65	1.00	5.17	6.92	1.16	.73	.95	2.29

Note. CO = cardiac output, TPR = total peripheral resistance, HR = heart rate, VC = ventricular contractility (preejection period × -1). Task performance reflects number of items correct out of 12 on the RAT.

†p<0.1 *$p < .05$. **$p < .01$. ***$p < .001$.

Challenge/threat.

Undermined competence hypothesis. Given that women who received BS feedback failed to exhibit cardiovascular responses consistent with lower task engagement than women who received non-sexist feedback, we addressed our favored hypothesis that they would instead exhibit cardiovascular responses consistent with relative threat. We thus tested for differences in cardiovascular challenge/threat responses as a function of condition, controlling for gender-role attitudes and gender identification. The main effect of sexism condition on the challenge threat index was significant, b=-0.53, $t(69)$=-2.50, p=.015, $\eta^2_{partial}$=0.08 (see Figure 1).[1,2] Parallel patterns emerged in separate analyses for CO (b=-0.52, $t(69)$=-1.83, p=.07, $\eta^2_{partial}$=0.05) and TPR (b=85.79, $t(69)$=2.74, p=.01, $\eta^2_{partial}$=0.10). Furthermore, this effect remained significant when performance on the practice (first) RAT was added to the regression model, b=-0.49, $t(68)$=-2.32, p=.02, $\eta^2_{partial}$=0.07. When women received BS feedback, they exhibited cardiovascular responses consistent with greater threat during the performance task compared to women in the non-sexist feedback condition.[3]

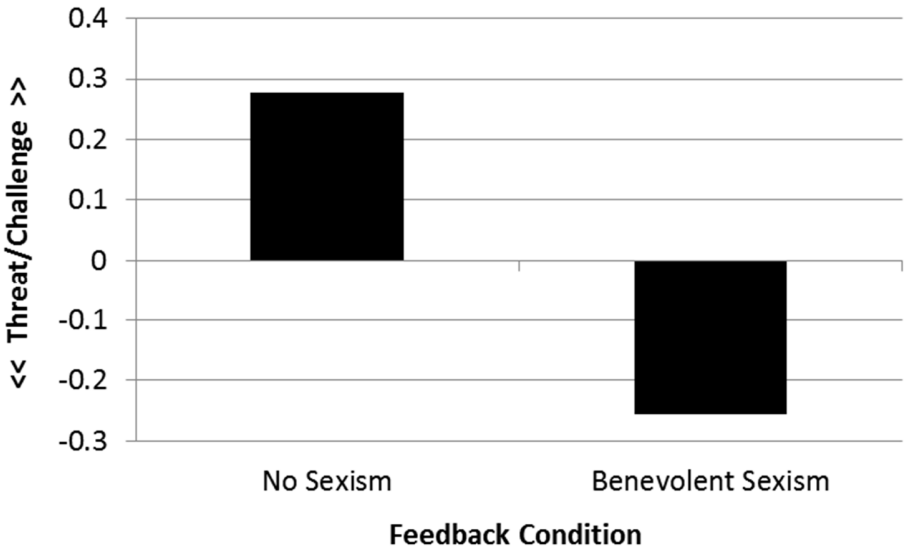

Figure 1. Feedback condition predicting cardiovascular challenge/threat responses (challenge/threat index), controlling for gender identification and sex-role attitudes. Lower scores reflect greater relative threat.

Moderating the responses to sexism. Consistent with our rationale for assessing the two covariates, higher endorsement of traditional gender roles was associated with greater challenge on the challenge/threat index (i.e., while performing in the gender-stereotypical domain of verbal performance), b=0.53, $t(69)$=3.66, p<.001, $\eta^2_{partial}$=0.16, as was stronger gender identification, b=0.20, $t(69)$=2.18, p=.033, $\eta^2_{partial}$=0.06. Separate analyses for CO and TPR yielded parallel patterns for gender role attitudes (CO: b=0.66, $t(69)$=3.40, p=.001, $\eta^2_{partial}$=0.14; TPR: b=-69.32, $t(69)$=-3.27, p=.002, $\eta^2_{partial}$=0.13) and gender identification (CO: b=.22, $t(69)$=1.82, p=.07, $\eta^2_{partial}$=0.05; TPR: b=-29.24, $t(69)$=-2.17, p=.03, $\eta^2_{partial}$=0.06).

Although our study was not designed to test the interaction between condition and the covariates, prior research has suggested that gender-role attitudes and gender identification may affect responses to sexism, albeit in different contexts than the current one (Schmader, 2002; McCoy & Major, 2002; Viki & Abrams, 2002; Russell & Triggs, 2004; Derks, Scheepers, Van Laar & Ellemers, 2011). We thus tested the interactions between feedback condition and gender-

role attitudes and gender identification in separate regression models, with the main effect of gender-role attitudes or gender identification remaining as a covariate (testing them together did not substantively affect results). For the challenge/threat index, an interaction between feedback condition and gender-role attitudes approached significance, b=-0.56, t(68)=-1.98, p=.052, $\eta^2_{partial}$=0.05 (see Figure 2). Testing simple effects of condition at +/– 1 *SD* from the mean of gender-role attitudes revealed that women with more traditional gender-role attitudes exhibited cardiovascular responses consistent with greater threat following BS compared to non-sexist feedback, b=-0.95, t(68)=-3.21, p=.002, $\eta^2_{partial}$=0.13. The simple effect of condition did not reach significance for women with less traditional gender-role attitudes, b=-0.13, t(68)=-0.44, p=0.66, $\eta^2_{partial}$=0.003. In sum, women with more traditional gender-role attitudes were more affected by BS feedback. Parallel interactions emerged in separate analyses for CO (b=-0.69, t(68)=-1.81, p=0.07, $\eta^2_{partial}$=0.05) and TPR (b=74.77, t(68)=1.79, p=0.08, $\eta^2_{partial}$=0.04).

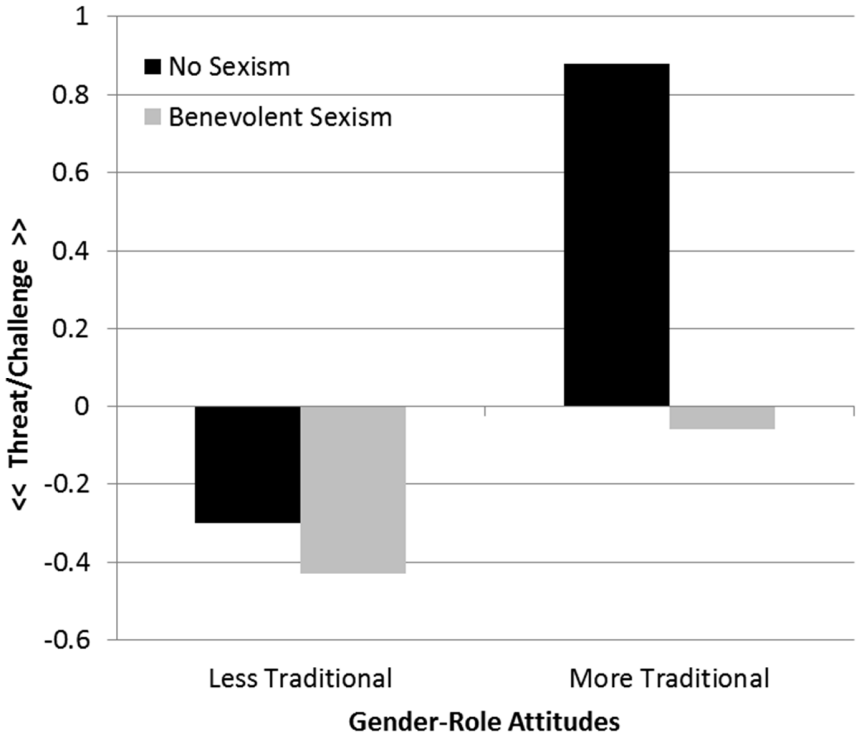

Figure 2. Feedback condition by sex-role attitudes interaction predicting cardiovascular challenge/threat responses (challenge/threat index), controlling for gender identification. Lower scores reflect greater relative threat.

The feedback condition by gender identification interaction was not significant for the challenge/threat index, b=0.03, t(68)=0.17, p=.86, $\eta^2_{partial}$<0.001 (CO, TPR ps > .40). The 3-way interaction between feedback condition, gender-role attitudes, and gender identification was also not significant, b=-0.02, t(65)=-0.06, p=.95, $\eta^2_{partial}$<0.001 (CO, TPR ps > .46).

Secondary Analyses: Performance

Test performance. A secondary question regarded whether feedback condition would affect RAT performance. Performance on the test (second) RAT was significantly correlated with performance on the practice RAT, b=.80, t(71)=2.95, p=.004, $\eta^2_{partial}$=.11. However, analyses did not reveal significant differences in performance by feedback condition, b=-0.53, t(70)=-.97, p=.34, $\eta^2_{partial}$=0.01, although the means were in the expected direction: Women exposed to BS feedback tended to perform worse (M=.56, SD=.94) than women who received non-sexist feedback (M=.84, SD=.95).

Self-reported performance. At the end of the study, participants were asked to reflect on various aspects of their test performance ("I tried hard during this task", "I did well on this task", "I am not skilled at this task", "I tried my best during this task", "I did not enjoy this task", "I would enjoy doing this task again", "This task was difficult", "This task was interesting"; 1=strongly disagree, 7=strongly agree). The only main effect of feedback condition that emerged was for perceived skill at the task (Table 2): Women who had received benevolent sexist feedback on the practice task reported believing that they were not skilled at the test task compared to women who had received non-sexist feedback, b=0.79, t(70)=2.22, p=.03, $\eta^2_{partial}$=0.07. Thus, despite showing no actual differences in performance on either the practice or

test performance tasks compared to women who received non-sexist women, women who had received benevolently sexist feedback felt less skilled following feedback that was intended to be supportive and encouraging.

Table 2
Model Coefficients for Self-Reported Performance

	Sexism Condition		Gender Identification		Gender Role Attitudes	
	b	*t*	*b*	*t*	*b*	*t*
Tried Hard	.16	.92	.02	.15	-.03	-.36
Did Well	.06	.20	-.14	-.62	-.21	-1.52
Not Skilled	.79	2.22*	-.09	-.35	.37	2.40*
Tried Best	.09	.37	.05	.30	.02	.24
Did Not Enjoy	.59	1.62	.24	.96	.08	.53
Would Enjoy Doing Again	-.53	-1.41	-.19	-.73	-.15	-.94
Was Difficult	.07	.27	-.17	-.98	.25	2.31*
Was Interesting	-.20	-.52	-.08	-.28	.01	.07

*Note. †p<0.1 *p<.05 **p<.01 *** p<.001*

Ensuring the Feedback was Benevolent

In order to be certain that the feedback used in our study was in fact being perceived as intended (benevolently rather than malevolently), we asked a new set of participants to evaluate our non-sexist, benevolent, and newly generated hostile sexist feedback.[4] Eighty-nine women who were not involved in Study 1 participated under the cover story that they would be asked to listen to a scenario involving a student and her teacher. In the scenario, the student performed poorly on a practice test and receives feedback from the instructor. Participants were randomly assigned to hear the non-sexist, benevolent sexist, or hostile sexist feedback. They were then asked to rate the instructor's feedback based on how helpful, supportive, patronizing, frustrating, upsetting, anger inducing, professional, and sexist it was, as well as how likely it was to make

the student feel smart and how likely she would be to perform better in the future using 7-point Likert scale responses (1=not at all, 7=very; Table 3).

Table 3
Feedback Ratings by Feedback Type

	Benevolent Sexist		Non-Sexist		Hostile Sexist		
Feedback Rating	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*
Helpful	3.10	1.40	3.63	1.73	1.72	1.10	13.82***
Supportive	4.23	1.63	5.37	1.54	1.66	.90	53.14***
Patronizing	3.77	1.56	3.17	1.56	5.59	1.76	16.39***
Frustrated	4.77	1.89	4.23	2.00	6.41	1.02	13.17***
Upset	4.50	1.76	3.60	1.73	6.28	1.28	21.07***
Angry	3.90	1.79	3.23	1.74	6.31	.93	32.38***
Happy	3.07	1.48	3.90	1.73	1.45	.78	23.43***
Motivated	3.47	1.85	4.67	1.84	3.00	2.30	5.45***
Friendly	4.77	1.57	5.73	1.17	3.24	1.79	19.92***
Professional	4.10	1.71	4.90	1.60	2.28	1.73	18.80***
Smart	4.60	1.22	4.37	1.27	4.55	1.35	.28
Sexist	3.40	2.14	2.33	1.30	6.41	1.09	52.64***
Will Do Better	3.77	1.48	3.83	1.76	3.72	2.02	.03

Note. †$p<0.1$ *$p<.05$ **$p<.01$ *** $p<.001$

Planned contrasts revealed that, as intended, the BS feedback was evaluatively different from hostile sexist feedback across all domains ($ps<0.001$), except for how smart the student seemed to raters and how likely the student was to do better the next time, which did not differ across any feedback type. In all differences, the BS feedback was rated more favorably than the hostile sexist feedback. This is consistent with extant research demonstrating that people who endorse attitudes or behave in a way consistent with BS ideologies are evaluated more positively than those who endorse and behave in line with hostile sexist ideologies.

Ratings of BS compared to non-sexist feedback also differed across some dimensions. Despite no differences between perceptions of feedback in terms of how helpful, patronizing, frustrating, anger inducing, and professional it seemed ($ps>.07$), BS feedback was rated as less

supportive, and more sexist than non-sexist feedback, and participants thought the student would feel less happy, less motivated, and more upset ($ps<0.03$). These differences are consistent with other work demonstrating that while BS is extensively more favorable compared to hostile sexist feedback, people may not see BS entirely magnanimously (Kilianski & Rudman, 1998). In sum, these additional data support that our manipulation had the intended effects.

Discussion

The aim of this study was to assess the impact of receiving supportive, yet benevolently sexist, feedback on cardiovascular responses during a subsequent test of reasoning ability. Previous research has demonstrated that women typically report feeling less efficacious, agentic and competent, and have lower task performance when they are the targets of benevolent sexism (Dardenne, Dumont, & Bollier, 2007; Jones et al., 2014). It has been argued that these cognitive evaluations and behavioral outcomes are due to learned helplessness, an inability to cope, and rumination about the implications of what was said, as women feel that they do not have the resources to combat the patronizing nature of the support that surreptitiously undervalues and undermines them (Vescio et al., 2005). While the existing work has suggested that these consequences are linked to deliberative contemplation about the BS remarks (Dardenne et al., 2007; Dardenne et al., 2013), we tested whether the negative consequences of BS can emerge more immediately, without opportunity to ruminate, and—importantly—while women are still engaged with a task related to the feedback they received (cf. Salomon et al., 2015). Specifically, we expected that women who receive benevolently sexist, yet supportive, feedback regarding their performance should feel their personal resources are unable to meet the demands of the situation, experience relative threat, and therefore exhibit the associated cardiovascular responses.

Consistent with our prediction, relative to women receiving non-sexist supportive feedback, the women in our study exhibited cardiovascular responses consistent with greater threat (lower CO/higher TPR) while completing a verbal test immediately following supportive BS feedback. We failed to find support for an alternative hypothesis that the negative consequences of BS could be better explained by women distancing or withdrawing from performance situations, as increases in HR and VC from baseline did not significantly differ as a function of the type of feedback women received. This suggests that although women were engaged and motivated to perform on the subsequent verbal reasoning test (consistent with task engagement), the BS feedback had undermined their perceived ability to reach that goal. These effects held regardless of individual differences in gender-role attitudes and gender identification, which had the potential to change the implications of the feedback.

These findings are novel in several ways. First, psychophysiological measures demonstrated that BS feedback can spontaneously negatively affect women's challenge/threat responses (reflecting resource/demand evaluations), even without undistracted opportunity for rumination. Second, the negative consequences of BS can persist into subsequent motivated performance situations, extending the consequences of a single sexist encounter into new experiences and tasks. Third, BS feedback can have negative implications even when it is intended to be supportive and in fact expresses positive expectations for performance. Although we did not find evidence to support differences in performance across BS and non-sexist feedback conditions (see Limitations and Future Directions below), women who had received BS feedback not only exhibited cardiovascular threat responses consistent with evaluations of personal resources not meeting situational demands, but also reported feeling less skilled despite not performing any worse than women who had received the non-sexist feedback. Over time,

women may disengage from activities that they feel less skilled or competent at. For instance, there has been a renewed interest in understanding the barriers women face that have led to disparities in pursuit of and persistence in STEM-related degrees and careers despite equal performance to men (Hango, 2013). The current findings suggest that if women are receiving feedback that is benevolently sexist, even if it is well-intentioned and supportive, they may nonetheless feel as though they cannot meet the demands of the field. These evaluations could become the difference between a young woman walking out of a math exam with an 80% grade believing she is not cut out for a STEM career, and another with the same mark signing up for the next level course.

Finally, these findings and implications highlight the novel extension this study represents for the biopsychosocial model of challenge/threat more broadly. The majority of the BPSC/T research on the cardiovascular consequences of prejudice among its targets has focused on explicitly negatively valenced situations (e.g., Flores, Chavez, Bolger, & Casad, 2019; Townsend, Major, Sawyer, & Mendes, 2010; Vick et al., 2008). Benevolent sexism is unique in that unlike hostile sexism or other overt forms of prejudice and discrimination, it is often deployed with good intentions, and often seen by the recipient in that light (Barreto & Ellemers, 2005; Bohner et al., 2010; Moya et al., 2007). Our findings suggest that the BPSC/T model is nonetheless sensitive enough to respond to the less explicit connotations of BS (e.g., paternalistic, undermining). Thus, this underscores how the BPSC/T model could be used to understand how potentially positive framing or encounters could lead to negative outcomes for both men and women.

Limitations and Future Directions

One limitation of this study is that it did not include self-reported assessments of personal resources or situational demands between the feedback and performance task. Although cardiovascular indexes of challenge and threat have been found to map onto explicit resource-demand evaluations in past work (e.g., Quigley, Barrett, & Weinstein, 2002; Schneider, 2008; Tomaka et al., 1993, 1997), the decision was made to omit such a self-report measure in this study for several reasons. First, we sought to avoid drawing attention to the link between the feedback and the subsequent task, which might have induced demand characteristics in our participants. Second, this study was intended to demonstrate that the consequences of benevolent sexism can emerge before women have the opportunity for conscious and deliberative appraisals of their performance or abilities. Thus, by moving directly from the feedback to the test task we eliminated this opportunity as much as possible. Instead, relying on cardiovascular methodology enabled us to assess the associated psychological processes *while* women were engaged in the task itself (Seery, 2013). Third, there is reason to believe that self-reports should not necessarily faithfully capture the psychological states and processes of interest in all contexts. For example, cardiovascular challenge/threat responses can correspond with relatively uncontrollable nonverbal behaviors rather than controllable ones (Weisbuch et al., 2009), suggesting that controllable responses such as self-report can be misleading (also see Blascovich, Mendes, & Seery, 2002). It is also possible for stimuli presented outside of conscious awareness to affect cardiovascular challenge/threat responses (Weisbuch-Remington et al., 2005), implicating the possibility of relatively automatic components in the evaluation process which may or may not be reflected in more controlled responses such as self-report (see Olson & Fazio, 2008). This could be particularly important in the context of BS feedback, which is frequently appraised as positive and well-intentioned despite also entailing underlying negative valence (Barreto &

Ellemers, 2005; Glick & Fiske, 1996). Including a self-reported measure of demand-resource evaluations could therefore have failed to capture any differences between conditions given the subtlety of the manipulation. Alternatively, a self-report measure before the test task could have obscured the results if it induced dissonance in participants by juxtaposing their positive appraisals against their less consciously accessible appraisals of their performance. And lastly, although performance differences did not emerge, women who received BS feedback reported feeling less skilled than women who did not. Although this was measured after the task and not immediately before, this assertion is consistent with low perceived resources and cardiovascular threat responses. Nonetheless, cardiovascular indexes, like any measure, are susceptible to error and bias (Behnke & Kaczmarek, 2018). Thus, future research should endeavor to demonstrate a clear link between resource-demand appraisals and cardiovascular responses following this type of feedback to further highlight the relationship between these effects.

Another limitation of this study is the failure to capture performance differences associated with cardiovascular responses or sexist feedback. Secondary analyses did not reveal significant differences in performance by condition, although women who received BS feedback tended to perform worse than those who did not. Both cardiovascular threat and benevolent sexism have been (separately) linked to performance costs (Blascovich et al, 2004; Dardenne et al., 2007; Jones et al., 2014). This was not the primary focus of the current study, as this particular task was chosen for its suitability for challenge/threat and amenability to a feedback-based manipulation, rather than its sensitivity to effects on performance quality. Other studies using this particular task have also failed to find performance-based differences following sexist feedback (Salomon et al., 2015).

That these performance deficits did not emerge in this study may be a function of the relatively novel methodology used in this paper. First, the test task was a verbal reasoning task, which is a positively stereotyped domain for women. It is possible that performance effects would have been more likely to emerge if women had been completing a task in a stereotype-threatened domain (e.g., math task) where performance would have been more stereotypically confirmatory (e.g., Park, Kondrak, Ward, & Streamer, 2018; Park, Young, Eastwick, Troisi, & Streamer, 2015). Another novel aspect of our design, which may account for the absence of performance differences, is that the feedback was intended to be supportive (vs. unsupportive or neutral). Many previous studies have relied on sexist feedback that is dismissive of a woman's ability more broadly (for examples see Dardenne et al., 2007; Solomon et al., 2015). It is possible, if not likely, that receiving supportive feedback is still better than receiving dismissive feedback. However, our findings still point to the negative consequences of even supportive feedback on demand-resource evaluations. Finally, the performance effects may have been obscured because of the test-retest nature of the performance paradigm whereby the test round was much easier than the practice round, resulting in better performance for everyone regardless of feedback. Again, because previous research has used feedback that was dismissive to women in general, not based on specific characteristics or abilities of the woman in the study, the test-retest nature of this study may highlight how additional research needs to differentiate the consequences of benevolent sexist feedback on novel versus repeated tasks and activities. Furthermore, it is possible that in this context, effect sizes are larger for psychological states like challenge/threat than performance, or that a 12-item RAT is not optimally sensitive for revealing performance differences. Thus, future work should investigate the effects of BS on challenge/threat and performance in different contexts, as well as over longer periods of time.

Additionally, future research should aim to better understand the personal attitudes and dispositions that may influence how people interpret and respond to sexism. Gender-role attitudes and gender identification were included as covariates a priori, as they have been shown to influence responses to sexism (Schmader, 2002; Russell et al., 2004). Though not planned when designing this study, the interaction between each covariate and feedback was tested. Lack of evidence for an interaction for gender identification suggests that women who identify less with their gender are not necessarily immune to BS feedback's negative impact. The marginal interaction for gender-role attitudes suggests that even relatively traditional women who may typically expect and appreciate stereotyped interactions (Townsend et al., 2010; Hammond & Overall, 2013) can still be harmed by BS. However, because this study was not designed to test these interactions, these conclusions should be considered speculative. Future research appropriately powered to test these effects could further explore the implications and reliability of these preliminary findings.

Likewise, although gender-role attitudes and gender identification were not found to reliably buffer against BS feedback in our study, future research should give consideration to other dispositional factors that may protect women from the negative consequences of BS feedback. Some women may be reassured by benevolent sexist ideologies (Cross, Overall, & Hammond, 2016). For instance, women who tend to endorse BS ideologies themselves may feel more threatened when men do not provide them with the feedback they expect (Fischer, 2006; Hammond & Overall, 2013; Jost & Kay, 2005; Sibley, Overall, & Duckitt, 2007). Individual differences in how generally efficacious people feel they are may also serve as a buffer against BS feedback. People high in self-esteem, not only typically believe in their abilities, but also believe that others see them favorably (Battistich, Solomon, & Delucchi, 1993; Baumeister, Tice,

& Hutton, 1989). They may therefore be better equipped to deflect the negative implications of BS feedback. Alternatively, people with low self-esteem feel most at ease with others when they receive feedback that confirms their self-views (Murray et al., 2005; Swann, Stein-Seroussi, & Giesler, 1992). Thus, they may feel more at ease with feedback that is superficially positive and supportive, but affirms their self-doubts more covertly, although we did not find any moderating effects of self-esteem in this study.

Lastly, another route for future research to consider is the characteristics of the individual providing the feedback and the relationship between that individual and the target. In our study, participants all received feedback from a male researcher whom they had not met. First, this raises the question as to whether the BS feedback would have the same impact coming from another woman. Although women can also endorse BS ideologies, and the theoretical implications of BS feedback are gender neutral, most research has focused on the impact of BS feedback delivered by a man. Thus, in order to gain a more ecologically sound understanding of the implications of BS in the real world, additional work should be done to see whether BS feedback is equally insidious when coming from a woman as a man. Second, people tend to wear rose-colored glasses when it comes to people whom they trust (Murray, Holmes, & Griffin, 1996), and they are more likely to make external attributions (e.g., “they’re from a different era”, “they didn’t mean it like *that*”) for negative behaviors (Kim, Dirks, Cooper, & Ferrin, 2006; Rempel, Ross, & Holmes, 2001; Tomlinson & Mayer, 2009). People may be more able or willing to downplay and disregard the implications of BS feedback if it comes from someone trusted versus an unknown other.

Conclusion

Cardiovascular measures of challenge/threat made it possible to assess the spontaneous negative consequences of benevolent sexism. Our findings provide further support for the insidious nature of benevolent sexism: Even when benevolently sexist feedback is intended as supportive and encouraging, it can undermine a woman's evaluation of her personal resources and skills, as evidenced by cardiovascular reactivity more akin to a threat state, during subsequent motivated performance tasks.

References

Abrams, D., Viki, G. T., Masser, B., & Bohner, G. (2003). Perceptions of stranger and acquaintance rape: The role of benevolent and hostile sexism in victim blame and rape proclivity. *Journal of Personality and Social Psychology, 84,* 111-125. doi: 10.1037/0022-3514.84.1.111

Barreto, M., & Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology, 35*, 633–642. doi:10.1002/ejsp.270

Battistich, V., Solomon, D., & Delucchi, K. (1993). Interaction processes and student outcomes in cooperative learning groups. *The Elementary School Journal, 94,* 19-32. doi:10.1086/461748

Baumeister, R. F., Tice, D. M., & Hutton, D. G. (1989). Self-presentational motivations and personality differences in self-esteem. *Journal of Personality, 57,* 547-579. doi:10.1111/j.1467-6494.1989.tb02384.x

Becker, J. C., Glick, P., Ilic, M., & Bohner, G. (2011). Damned if she does, damned if she doesn't: Consequences of accepting versus confronting patronizing help for the female target and male actor. *European Journal of Social Psychology*, *41*, 761–773. doi:10.1002/ejsp.823

Becker, J. C., & Wright, S. C. (2011). Yet another dark side of chivalry: Benevolent sexism undermines and hostile sexism motivates collective action for social change. *Journal of Personality and Social Psychology, 101*, 62–77. doi:10.1037/a0022615

Behnke, M., & Kaczmarek, L. D. (2018). Successful performance and cardiovascular markers of challenge and threat: A meta-analysis. *International Journal of Psychophysiology, 130*, 73-79. doi: 10.1016/j.ijpsycho.2018.04.007

Blascovich, J. (2008). Challenge and threat. In A. J. Elliot (Ed.), Handbook of Approach and Avoidance Motivation (pp. 431–445). New York: Psychology Press.

Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social "facilitation" as challenge and threat. *Journal of Personality and Social Psychology, 77*, 68-77. doi: 10.1037/0022-3514.77.1.68

Blascovich, J., Mendes, W. B., & Seery, M. D. (2002). Intergroup encounters and threat: A multi-method approach. In D. M. Mackie & E. R. Smith (Eds.), *From prejudice to intergroup emotions: Differentiated reactions to social groups* (pp. 89-109). New York: Psychology Press.

Blascovich, J., Seery, M. D., Mugridge, C. A., Norris, K., & Weisbuch, M. (2004). Predicting athletic performance from cardiovascular indexes of challenge and threat. *Journal of Experimental Social Psychology, 40,* 683-688. doi:10.1016/j.jesp.2003.10.007

Blascovich, J., & Tomaka, J. (1996). The Biopsychosocial model of arousal regulation. *Advances in Experimental Social Psychology*, *28*, 1–51. doi:10.1016/s0065-2601(08)60235-x

Bohner, G., Ahlborn, K., & Steiner, R. (2010). How sexy are sexist men? Women's perception of male response profiles in the Ambivalent Sexism Inventory. *Sex Roles, 62,* 568-582. doi:10.1007/s11199-009-9665-x

Brownley, K. A., Hurwitz, B. E., & Schneiderman, N. (2000). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 224-264). New York: Cambridge University Press.

Cross, E. J., Overall, N. C., & Hammond, M. D. (2016). Perceiving partners to endorse benevolent sexism attenuates highly anxious women's negative reactions to conflict. *Personality and Social Psychology Bulletin, 42,* 923-940. doi:10.1177/0146167216647933

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology*, *93*, 764–779. doi:10.1037/0022-3514.93.5.764

Dardenne, B., Dumont, M., Sarlet, M., Phillips, C., Balteau, E., Degueldre, C., … Collette, F. (2013). Benevolent sexism alters executive brain responses. *NeuroReport*, *24*, 572–577. doi:10.1097/wnr.0b013e3283625b5b

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615–1628. doi:10.1177/014616702237644

de Wit, F. R. C., Scheepers, D., & Jehn, K. A. (2012). Cardiovascular reactivity and resistance to opposing viewpoints during intragroup conflict. *Psychophysiology, 49*, 1691-1699. doi:10.1111/j.1469-8986.2012.01456.x

Derks, B., Scheepers, D., Van Laar, C., & Ellemers, N. (2011). The threat vs. Challenge of car parking for women: How self- and group affirmation affect cardiovascular responses. *Journal of Experimental Social Psychology, 47*, 178–183. doi:10.1016/j.jesp.2010.08.016

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review, 96*, 84-100. doi: 10.1037/0033-295X.96.1.84

Dumont, M., Sarlet, M., & Dardenne, B. (2010). Be too kind to a woman, she'll feel incompetent: Benevolent sexism shifts self-construal and autobiographical memories toward incompetence. *Sex Roles*, 62, 545-553. doi: 10.1007/s11199-008-9582-4

Fischer, A. R. (2006). Women's benevolent sexism as reaction to hostility. *Psychology of Women Quarterly, 30,* 410-416.

Flores, A. J., Chavez, T. A., Bolger, N., & Casad, B. J. (2019). Cardiovascular and self-regulatory consequences of SES-based social identity threat. *Personality and Social Psychology Bulletin, 45*, 700-714. doi: 10.1177/0146167218795157

Fowles, D. C., Fisher, A. E., & Tranel, D. T. (1982). The heart beats to reward: The effect of monetary incentive on heart rate. *Psychophysiology, 19*, 506-513. doi: 10.1111/j.1469-8986.1982.tb02577.x

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491–512. doi:10.1037//0022-3514.70.3.491

Glick, P., & Fiske, S. T. (1997). Hostile and benevolent sexism. *Psychology of Women Quarterly*, *21*, 119–135. doi:10.1111/j.1471-6402.1997.tb00104.x

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *56*, 109–118. doi:10.1037/0003-066x.56.2.109

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., … López, W. L. (2000). Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, *79*, 763–775. doi:10.1037//0022-3514.79.5.763

Glynn, L. M., Christenfeld, N., & Gerin, W. (2002). The role of rumination in recovery from reactivity: Cardiovascular consequences of emotional states. *Psychosomatic Medicine, 64,* 714-726.

Hammond, M. D., & Overall, N. C. (2013). When relationships do not live up to benevolent ideals: Women's benevolent sexism and sensitivity to relationship problems. *European Journal of Social Psychology*, *43*, 212–223. doi:10.1002/ejsp.1939

Hammond, M. D., & Overall, N. C. (2015). Benevolent sexism and support of romantic partner's goals: Undermining women's competence while fulfilling men's intimacy needs. *Personality and Social Psychology Bulletin*, *41*, 1180–1194. doi:10.1177/0146167215593492

Hango, D. (2013). Gender differences in science, technology, engineering, mathematics, and computer science (STEM) programs at university. Statistics Canada: Insights on Canadian Society, Catalogue no. 75-006-X, ISSN 2291-0840.

Hase, A., O'Brien, J., Moore, L. J., & Freeman, P. (in press). The relationship between challenge and threat states and performance: A systematic review. *Sport, Exercise, and Performance Psychology*. doi: 10.1037/spy0000132

Jones, K., Stewart, K., King, E., Botsford Morgan, W., Gilrane, V., & Hylton, K. (2014). Negative consequence of benevolent sexism on efficacy and performance. *Gender in Management: An International Journal*, *29*, 171–189. doi:10.1108/gm-07-2013-0086

Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, *88*, 498–509. doi:10.1037/0022-3514.88.3.498

Kehn, A., & Ruthig, J. C. (2013). Perceptions of gender discrimination across six decades: The moderating roles of gender and age. *Sex Roles, 69,* 289-296. doi:10.1007/s11199-013-0303-2. re

Key, B. L., Campbell, T. S., Bacon, S. L., & Gerin, W. (2008). The influence of trait and state rumination on cardiovascular recovery from a negative emotional stressor. *Journal of Behavioral Medicine, 31,* 237-248.

Kilianski, S. E., & Rudman, L. A. (1998). Wanting it both ways: Do women approve of benevolent sexism? *Sex Roles, 39,* 333-352.

Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence-vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes, 99,* 49-65. doi:10.1016/j.obhdp.2005.07.002

King, E. B., Botsford, W., Hebl, M. R., Kazama, S., Dawson, J. F., & Perkins, A. (2012). Benevolent sexism at work: Gender differences in the distribution of challenging developmental experiences. *Journal of Management, 38,* 1835-1866. doi: 10.1177/0149206310365902

Le, P. Q., Saltsman, T. L., Seery, M. D., Ward, D. E., Kondrak, C. L., & Lamarche, V. M. (2019). When a small self means manageable obstacles: Spontaneous self-distancing predicts divergent effects of awe during a subsequent performance stressor. *Journal of Experimental Social Psychology, 80*, 59-66. doi: 10.1016/j.jesp.2018.07.010

Lee, T. L., Fiske, S. T., Glick, P., & Chen, Z. (2010). Ambivalent sexism in close relationships: (hostile) power and (benevolent) romance shape relationship ideals. *Sex Roles*, *62*, 583–601. doi:10.1007/s11199-010-9770-x

Llabre, M. M., Spitzer, S. B., Saab, P. G., Ironson, G. H., & Schneiderman, N. (1991). The reliability and specificity of delta versus residualized change as measures of cardiovascular reactivity to behavioral challenges. *Psychophysiology, 28*, 701-711. doi: 10.1111/j.1469-8986.1991.tb01017.x

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of One's social identity. *Personality and Social Psychology Bulletin*, *18*, 302–318. doi:10.1177/0146167292183006

Lupien, S. P., Seery, M. D., & Almonte, J. L. (2012). Unstable high self-esteem and the eliciting conditions of self-doubt. *Journal of Experimental Social Psychology, 48*, 762-765. doi: 10.1016/j.jesp.2012.01.009

McCoy, S. K., & Major, B. (2003). Group identification moderates emotional responses to perceived prejudice. *Personality and Social Psychology Bulletin*, *29*, 1005–1017. doi:10.1177/0146167203253466

McFarlin, D. B., & Blascovich, J. (1984). On the Remote Associates Test (RAT) as an alternative to illusory performance feedback: A methodological note. *Basic and Applied Social Psychology, 5*, 223-229. doi: 10.1207/s15324834basp0503_5

Moore, L. J., Vine, S. J., Wilson, M. R., & Freeman, P. (2012). The effect of challenge and threat states on performance: An examination of potential mechanisms. *Psychophysiology, 49*, 1417-1425. doi: 10.1111/j.1469-8986.2012.01449.x

Moore, L. J., Vine, S. J., Wilson, M. R., & Freeman, P. (2014). Examining the antecedents of challenge and threat states: The influence of perceived required effort and support availability. *International Journal of Psychophysiology*, *93*, 267-273. doi: 10.1016/j.ijpsycho.2014.05.009

Moore, L. J., Wilson, M. R., Vine, S. J., Coussens, A. H., & Freeman, P. (2013). Champ or chump?: Challenge and threat states during pressurized competition. *Journal of Sport & Exercise Psychology, 35*, 551-562. doi: 10.1123/jsep.35.6.551

Moss-Racusin, C.A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handlesman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109,* 16474-16479. doi:10.1073/pnas.121126109.

Moya, M., Glick, P., Exposito, F., de Lemus, S., & Hart, J. (2007). It's for your own good: benevolent sexism and women's reactions to protectively justified restrictions. *Personality and Social Psychology Bulletin, 33,* 1421–1434. doi:10.1177/0146167207304790

Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality and Social Psychology, 70,* 79-98. doi:10.1037/0022-3514.70.1.79

Murray, S. L., Rose, P., Holmes, J. G., Derrick, J., Podchaski, E. J., Bellavia, G., & Griffin, D. W. (2005). Putting the partner within reach: a dyadic perspective on felt security in close relationships. *Journal of Personality and Social Psychology, 88*, 327. doi:10.1037/0022-3514.88.2.327

Olson, M. A., & Fazio, R. H. (2008). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 19-63). New York: Psychology Press.

Overall, N. C., Sibley, C. G., & Tan, R. (2011). The costs and benefits of sexism: Resistance to influence during relationship conflict. *Journal of Personality and Social Psychology*, *101*, 271–290. doi:10.1037/a0022727

Papillo, J. F., & Shapiro, D. (1990). The cardiovascular system. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements.* (pp. 456-512). New York: Cambridge University Press.

Park, L. E., Kondrak, C. L., Ward, D. E., & Streamer, L. (2018). Positive feedback from male authority figures boosts women's math outcomes. *Personality and Social Psychology Bulletin, 44,* 359-383. doi: 10.1177/0146167217741312

Park, L. E., Young, A. F., Eastwick, P. W., Troisi, J. D., & Streamer, L. (2016). Desirable but not smart: Preference for smarter romantic partners impairs women's STEM outcomes. *Journal of Applied Social Psychology, 46*, 158-179. doi: doi.org/10.1111/jasp.12354

Quigley, K. S., Barrett, L. F., & Weinstein, S. (2002). Cardiovascular patterns associated with threat and challenge appraisals: A within-subjects analysis. *Psychophysiology, 39*, 292-302. doi: 10.1017/s0048577201393046

Rempel, J. K., Ross, M., & Holmes, J. G. (2001). Trust and communicated attributions in close relationships. *Journal of Personality and Social Psychology, 81,* 57-64. doi:I0.1037//0022-3514.8I.1.57

Russell, B. L., & Trigg, K. Y. (2004). Tolerance of sexual harassment: An examination of gender differences, ambivalent sexism, social dominance, and gender roles. *Sex Roles*, *50*, 565–573. doi:10.1023/b:sers.0000023075.32252.fd

Salomon, K., Burgess, K. D., & Bosson, J. K. (2015). Flash fire and slow burn: Women's cardiovascular reactivity and recovery following hostile and benevolent sexism. *Journal of Experimental Psychology: General*, *144*, 469–479. doi:10.1037/xge0000061

Saltsman, T. L., Seery, M. D., Kondrak, C. L., Lamarche, V. M., & Streamer, L. (2019). Too many fish in the sea: A motivational examination of the choice overload experience. *Biological Psychology, 145*, 17-30. doi: 10.1016/j.biopsycho.2019.03.010

Scheepers, D. (2009). Turning social identity threat into challenge: Status stability and cardiovascular reactivity during inter-group competition. *Journal of Experimental Social Psychology*, *45*, 228–233. doi:10.1016/j.jesp.2008.09.011

Scheepers, D., de Wit, F., Ellemers, N., & Sassenberg, K. (2012). Social power makes the heart work more efficiently: Evidence from cardiovascular markers of challenge and threat. *Journal of Experimental Social Psychology, 48*, 371-374. doi: 10.1016/j.jesp.2011.06.014

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 194–201. doi:10.1006/jesp.2001.1500

Schneider, T. R. (2008). Evaluations of stressful transactions: What's in an appraisal? *Stress and Health, 24*, 151-158. doi: 10.1002/smi.1176

Seery, M. D. (2011). Challenge or threat? Cardiovascular indexes of resilience and vulnerability to potential stress in humans. *Neuroscience & Biobehavioral Reviews*, *35*, 1603–1610. doi:10.1016/j.neubiorev.2011.03.003

Seery, M. D. (2013). The Biopsychosocial model of challenge and threat: Using the heart to measure the mind. *Social and Personality Psychology Compass*, *7*, 637–653. doi:10.1111/spc3.12052

Seery, M. D., Kondrak, C. L., Streamer, L., Saltsman, T., & Lamarche, V. M. (2016). Preejection period can be calculated using R peak instead of Q. *Psychophysiology, 53*, 1232-1240. doi: 10.1111/psyp.12657s

Seery, M. D., & Quinton, W. J. (2016). Understanding resilience: From negative life events to everyday stressors. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 54, pp. 181-245). Cambridge, MA: Academic Press.

Seery, M. D., Weisbuch, M., & Blascovich, J. (2009). Something to gain, something to lose: The cardiovascular consequences of outcome framing. *International Journal of Psychophysiology, 73*, 308-312. doi:10.1016/j.ijpsycho.2009.05.006

Seery, M. D., Blascovich, J., Weisbuch, M., & Vick, S. B. (2004). The relationship between self-esteem level, self-esteem stability and cardiovascular reactions to performance feedback. *Journal of Personality and Social Psychology, 87*, 133–145. doi: 10.1037/0022-3514.87.1.133

Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology, 87,*38-56. doi:10.1037/0022-3514.87.1.38

Shimizu, M., Seery, M. D., Weisbuch, M., & Lupien, S. P. (2011). Trait social anxiety and physiological activation: Cardiovascular threat during social interaction. *Personality and Social Psychology Bulletin, 37*, 94-106. doi: 10.1177/0146167210391674

Shnabel, N., Bar-Anan, Y., Kende, A., Bareket, O., & Lazar, Y. (2016). Help to perpetuate traditional gender roles: Benevolent sexism increases engagement in dependency-oriented cross-gender helping. *Journal of Personality and Social Psychology, 110*, 55–75. doi:10.1037/pspi0000037

Sibley, C. G., Overall, N. C., & Duckitt, J. (2007). When women become more hostilely sexist toward their gender: The system-justifying effect of benevolent sexism. *Sex Roles, 57*, 743. doi:10.1007/s11199-007-9306-1

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles, 50*, 241–252. doi:10.1023/b:sers.0000015555.40976.e6

Streamer, L., Seery, M. D., Kondrak, C. L., Lamarche, V. M., & Saltsman, T. L. (2017). Not I, but she: The beneficial effects of self-distancing on challenge/threat cardiovascular responses. *Journal of Experimental Social Psychology, 70*, 235-241. doi: 10.1016/j.jesp.2016.11.008

Swann, W. B., Stein-Seroussi, A., & Giesler, R. B. (1992). Why people self-verify. *Journal of Personality and Social Psychology, 62,* 392.

Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology, 65*, 248-260. doi: 10.1037/0022-3514.65.2.248

Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology, 73*, 63-72. doi: 10.1037/0022-3514.73.1.63

Tomlinson, E. C., & Mryer, R. C. (2009). The role of causal attribution dimensions in trust repair. *Academy of Management Review, 34,* 85-104.

Townsend, S. S. M., Major, B., Sawyer, P. J., & Mendes, W. B. (2010). Can the absence of prejudice be more threatening than its presence? It depends on one's worldview. *Journal of Personality and Social Psychology, 99*, 933–947. doi:10.1037/a0020434

Tranel, D. T., Fisher, A. E., & Fowles, D. C. (1982). Magnitude of incentive effects on heart rate. *Psychophysiology, 19*, 514-519. doi: 10.1111/j.1469-8986.1982.tb02578.x

Turner, M. J., Jones, M. V., Sheffield, D., & Cross, S. L. (2012). Cardiovascular indices of challenge and threat states predict competitive performance. *International Journal of Psychophysiology*, *86*, 48-57. doi: 10.1016/j.ijpsycho.2012.08.004

Turner, M. J., Jones, M. V., Sheffield, D., Barker, J. B., & Coffee, P. (2014). Manipulating cardiovascular indices of challenge and threat using resource appraisals. *International Journal of Psychophysiology, 94*, 9-18. doi: 10.1016/j.ijpsycho.2014.07.004

Turner, M. J., Jones, M. V., Sheffield, D., Slater, M. J., Barker, J. B., & Bell, J. J. (2013). Who thrives under pressure? Predicting the performance of elite academy cricketers using the cardiovascular indicators of challenge and threat states. *Journal of Sport & Exercise Psychology*, *35*, 387-397.

Van Yperen, N. W., & Buunk, B. P. (1991). Sex-role attitudes, social comparison, and satisfaction with relationships. *Social Psychology Quarterly*, *54*, 169. doi:10.2307/2786934

Vescio, T. K., Gervais, S. J., Snyder, M., & Hoover, A. (2005). Power and the creation of patronizing environments: The stereotype-based behaviors of the powerful and their effects on female performance in masculine domains. *Journal of Personality and Social Psychology*, *88*, 658–672. doi:10.1037/0022-3514.88.4.658

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*, 624–630. doi:10.1016/j.jesp.2007.02.007

Viki, G. T., & Abrams, D. (2002). But she was unfaithful: Benevolent sexism and reactions to rape victims who violate traditional gender role expectations. *Sex Roles*, *47*, 289–293. doi:10.1023/A:1021342912248

Viki, G. T., & Abrams, D. (2003). Infra-humanization: Ambivalent sexism and the attribution of primary and secondary emotions to women. *Journal of Experimental Social Psychology*, *39*, 492–499. doi:10.1016/s0022-1031(03)00031-3

Weisbuch, M., Seery, M., Ambady, N., & Blascovich, J. (2009). On the correspondence between physiological and nonverbal responses: Nonverbal behavior accompanying challenge and threat. *Journal of Nonverbal Behavior, 33*, 141-148. doi: 10.1007/s10919-008-0064-8

Weisbuch-Remington, M., Mendes, W. B., Seery, M. D., & Blascovich, J. (2005). The nonconscious influence of religious symbols in motivated performance situations. *Personality and Social Psychology Bulletin, 31*, 1203-1216. doi: 10.1177/0146167205274448

Williams, S., Cumming, J. & Balanos, G. (2010). The use of imagery to manipulate challenge and threat appraisal states in athletes. *Journal of Sport & Exercise Psychology, 32,* 339-358. doi: 10.1123/jsep.32.3.339

Notes

[1] When gender identification and sex-role attitudes were removed from analyses, the main effect of condition became marginally significant, b=-0.40, t(71)=-1.74, p=0.086, $\eta^2_{partial}$=0.04.

[2] The decision to look at the first two minutes of reactivity during the test RAT was made a priori, consistent with previous research using the task (e.g., Seery et al., 2004; Lupien et al., 2012). However, the pattern of results remained consistent whether we tested the first minute individually (b=-.45, t(68)=-1.96, p=.054, $\eta^2_{partial}$=0.05) or the mean across all three minutes (b=-.50, t(69)=-2.33, p=.023, $\eta^2_{partial}$=0.07).

[3] Self-esteem was assessed using the Rosenberg Self-Esteem Scale (1965) as a filler task during the initial demographic assessment and lab acclimation period. Because no effects of self-esteem were predicted a priori, they were not included in the initial analyses. However, in order to test whether individual differences in self-efficacy would have better accounted for or buffered against the effects of feedback on cardiovascular reactivity, we re-ran the analyses adding self-esteem to the model. The effect of feedback condition, b=-.53, t(68)=-2.39, p=.02, $\eta^2_{partial}$=0.08, gender identification, b=.20, t(68)=2.17, p=.03, $\eta^2_{partial}$=0.06, and sex-role attitudes, b=.53, t(68)=3.57, p=.001, $\eta^2_{partial}$=0.16, remained significant when controlling for self-esteem. However, the effect of self-esteem was not significant, b=.01, t(68)=.11, p=.91, $\eta^2_{partial}$<0.001. Furthermore, the feedback condition by self-esteem interaction was not significant, b=-.05, t(67)=-.22, p=.82, $\eta^2_{partial}$<0.001.

[4] Those in the hostile sexist feedback condition heard the following, delivered by the same man who recorded the other two conditions, with a tone that was not supportive: You seem ***very smart for a girl*** because your answers showed a lot of creativity, but ***you're still going to struggle*** with the next set of ***questions because girls don't do very well on this test.*** I guess just do your best to come up with answers and ***don't let your nerves get the best of you.***

Appendix A

Gender Identification Scale

Answer each item based on your gender (i.e., if you are female, answer the items based on being a woman, and vice versa for men). There are no right or wrong answers so please be as honest as possible in your responses.

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

1. Overall, being a woman/man has very little to do with how I feel about myself.
2. Being a woman/man is an important reflection of who I am.
3. Being a woman/man is unimportant to my sense of what kind of person I am.
4. In general, being a woman/man is an important part of my self-image.

Gender-Role Attitudes

Please indicate how much you agree with the following statements.

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

1. Everything should be geared toward assuring that as many women as men are employed in all professions.
2. It looks worse for a woman than for a man to be drunk.
3. A woman who has children should be a mother before all else.
4. I think it is wrong that a man cannot as easily take on the name of his wife when he marries - if he should want to, that is.
5. I think it is wrong that the media (newspapers, television) pay more attention to men's sports than to women's sports.
6. It is not appropriate for a woman to tell dirty jokes.
7. It is ridiculous for a woman to help a man put on his coat.
8. It is acceptable for a woman to have a career, but marriage and family should come first.
9. I think feminism is an important cause.
10. A woman must insist that the domestic chores be divided equally between the two spouses.
11. A woman should not attempt to take on all kinds of typically male tasks.
12. A man who easily becomes emotional and breaks into tears is a softie.
13. It annoys me that men are better off than women in all possible respects.

14. As long as men dominate in our society, it's not possible for women to be really happy.
15. From the beginning of a relationship, a woman has to be careful that she isn't pushed into the traditional female role.
16. The development of traditional gender roles are a logical consequence of getting married and having children.
17. It is best to maintain more or less traditional gender roles in a relationship.

**Denotes items that were reverse-scored.*

Self-Reported Evaluation of Test RAT

Please rate how much you agree or disagree with each of the following statements.

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

1. I tried hard during this task.
2. I did well on this task.
3. I am not skilled at this task.
4. I tried my best during this task.
5. I did not enjoy this task.
6. I would enjoy doing this task again.
7. This task was difficult.
8. This task is interesting.

Perceptions of Feedback (Manipulation Check Study)

[Instructions]

In this study we are interested in better understanding student-teacher interactions in the classroom. You will be asked to read and listen to a scenario involving a student and his or her teacher and will then be asked to answer several questions about the situation, the student and the teacher.

[Click to continue]

SCENARIO # 23:

Kelly is an English major enrolled in an advanced language seminar with Dr. Robert Marron.

[Click to continue]

To assess his class' language skills, Dr. Marron sets up a practice test and then gives each student personalized feedback.

Kelly scored 3/12 on the practice test.

[Click to continue]

Dr. Marron gave Kelly the following feedback in person:

[-all-]

OK, so it looks like you struggled with this test so far. But the first set of questions were just practice questions, so the next set that you'll answer will be what you are evaluated on.

[-students randomly assigned to one of the following-]

BS Feedback: *You seem like a **very smart girl** because your answers showed a lot of creativity. I know it's hard not to get **emotional** during this type of test, but I'm sure you'll do well on the next set of questions as long as you **don't let your nerves get the best of you**.*

[-or-]

Non-Sexist Feedback: *You seem like a **very smart person** because your answers showed a lot of creativity. I know it's hard to **come up with answers** during this type of test, but I'm sure you'll do well on the next set of questions as long as you **continue to think outside of the box**.*

[-or-]

HS Feedback: You seem ***very smart for a girl*** because your answers showed a lot of creativity, but ***you're still going to struggle*** with the next set of ***questions because girls don't do very well on this test.*** I guess just do your best to come up with answers and ***don't let your nerves get the best of you.***

[Click to Continue]

Now we would like to ask you some questions about the interaction between Kelly and Dr. Robert Marron.

1	2	3	4	5	6	7
Not at all						Very

1. How helpful was Dr. Marron's feedback to Kelly?
2. How supportive was Dr. Marron's feedback to Kelly?
3. How patronizing was Dr. Marron's feedback to Kelly?
4. How frustrated would Kelly have been by Dr. Marron's feedback?
5. How upset would Kelly have been by Dr. Marron's feedback?
6. How angry would Kelly have been after Dr. Marron's feedback?
7. How happy would Kelly have been after Dr. Marron's feedback?
8. How motivated would Kelly have been after Dr. Marron's feedback?
9. How friendly is Dr. Marron?
10. How professional is Dr. Marron?
11. How smart is Kelly?
12. How sexist is Dr. Marron?
13. Do you think Kelly will do better on the next set of questions?

Appendix B

Participants heard the following pre-recorded instructions throughout the study.

Baseline Instructions:

Welcome to the social psychophysiology laboratory. We would like to thank you for your participation today. Please try to move as little as possible for the rest of the experiment, although you may move when necessary. When possible, rest your hands on top of the lap tray. Please do not touch the keyboard unless you are instructed to do so. Before the study begins, we need to calibrate our physiological equipment. This occurs entirely in the experimenter's control room, so it will seem to you like nothing is happening. All we need you to do is sit quietly. This will take a few minutes. For the next few minutes, please sit quietly and relax until the experimenter tells you it is time to continue.

Practice RAT Instructions:

In this study, we are interested in physiological responses during tests of academic aptitude and ability. You are about to take the Remote Associates Test. Higher scores on this test predicts academic and future career success. People who receive high scores on this test are more likely to do well in college and be accepted to graduate and professional programs.

This test is made up of 12 items. Each item will appear on the screen for 15 seconds. You must say your answers aloud so that the experimenter can record them. You will be presented with a number of three word-groups from which you

will need to generate the single word that links all three words together. It is very important that you make a guess for each set of words, even if you can only think of an answer that applies to one or two of the words. Press the spacebar to continue.

In this example, the following prompt words appear on the screen: sea, home, and stomach. The single word that links these three words together is "sick", as in "seasick", "homesick", and "sick to your stomach". For this example, you would say the word "sick" aloud. Sometimes the answer goes with prompt words to form a phrase, like "seasick" and "homesick". However, sometimes the answer is only conceptually linked to prompt words, like in "sick to your stomach". The test item may include either or both kinds of these relationships. Once the test starts, you will have only 15 seconds to answer each of the 12 items. Once the 15 seconds have passed, the computer will automatically move on to the next item. You cannot go back, so it is important that you say the answer aloud if you think you have one. The experimenter will only record the last answer that you give for each item, and you must respond before 15 seconds are up. The experimenter cannot tell you if you have answered an item correctly or what the correct answer is. Before the computer moves on to the next item, it will briefly show the words "next item" displayed in red in the middle of the screen. Press the spacebar to see the next item.

In this example, the following prompt words appear on the screen: milk, farm, and bell. The correct answer is "cow". "Milk" comes from cows, cows live on "farm(s)", and a cow "bell" is a type of bell. Answers can be related to prompt words in many different ways, as can be seen in this example. Remember, it is important to make a guess out loud even if you're not 100% sure your answer is correct.

[After ~3 minutes:]

Time is up.

Sexism Manipulation:

Experimenter (via the intercom):

Okay, you have now completed the first set of questions and got [number correct**] out of 12 correct. Before you complete the next set of questions, our lead researcher in charge of the lab is going to review your performance so far and provide you with some feedback to help you with the next part of this test. While the lead researcher reviews your performance we will also need to calibrate some of our equipment.**

This will take a few minutes. Please sit back and relax. I will let you know when we are ready to continue.

[After 5 minutes:]

Okay, now I'm connecting your intercom to the lead researcher who will provide you with feedback on your performance so far.

Recorded Feedback:

BS Feedback:

OK, so it looks like you struggled with this test so far. But the first set of questions were just practice questions, so the next set that you will answer will be what you are evaluated on. You seem like a very smart girl because your answers showed a lot of creativity. I know it's hard not to get emotional during this type of test, but I'm sure you'll do well on the next set of questions as long as you don't let your nerves get the best of you.

----- OR -----

Non-Sexist Feedback:

OK, so it looks like you struggled with this test so far. But the first set of questions were just practice questions, so the next set that you will answer will be what you are evaluated on. You seem like a very smart person because your answers showed a lot of creativity. I know it's hard to come up with answers during this type of test, but I'm sure you'll do well on the next set of questions as long as you continue to think outside of the box.

Test RAT Instructions:

Now you are about to take a different version of the reasoning-ability test that you just took. The items will be different, but the format will be exactly the same. The first test was a practice test; only the results from this second test will be recorded. We will be videotaping your performance and members of the research team will analyze your responses. We will use your results from this test to determine your level of intelligence and academic ability. Remember to speak loud enough so that the experimenter can hear your answers.

[After ~3 minutes:]

Time is up.