Words Like Weapons: Labeling Women As Emotional During a Disagreement Negatively Affects the Perceived Legitimacy of Their Arguments

Teresa J. Frasca¹, Emily A. Leskinen², and Leah R. Warner²

Abstract
With one in eight Americans thinking women are too emotional to be in politics (Carnevale et al., 2019), being labeled as emotional during a disagreement may activate stereotypes about a woman’s irrationality and affect how legitimate people perceive her arguments to be. We experimentally tested the effects of such labels. In Study 1 (N = 86), participants who read a vignette where a woman (versus a man) was told to “calm down” during a disagreement, saw her argument as significantly less legitimate. Perceived emotionality mediated the relation between condition and perceived legitimacy. Study 2 replicated this finding (N = 126) with different vignettes where the character was explicitly labeled as “emotional.” Using video vignettes in Study 3 (N = 251), we failed to replicate the results observed in Studies 1 and 2. We hope practitioners use these studies to increase awareness of how stereotype-laden labels can delegitimize women’s arguments, particularly when heard via writing (e.g., via email, text, or instant messaging) rather than when observed. This work may motivate observers to challenge the use of delegitimizing labels, so that women’s claims can be judged based on the soundness of their arguments, rather than stereotypes about their ability to think rationally. Additional online materials for this article are available on PWQ’s website at http://journals.sagepub.com/doi/suppl/10.1177/03616843221123745

Keywords
gender, emotion, labels, legitimacy, stereotypes

In 2011, Britain’s Prime Minister, David Cameron, told Angela Eagle, a woman opposition lawmaker, to “calm down, dear” in response to a typically heated argument in Parliament (BBC News, 2011). In 2016, presidential candidate Donald Trump labeled candidate Hillary Clinton as “a totally unhinged person” who is a “dangerous liar” (Del Real, 2016). More recently, a Twitter user claimed that Alexandria Ocasio-Cortez’s “frequent crying only reinforces the stereotype that women are too emotional for politics” (Prince, 2019). These instances represent a common reaction to—and political strategy against—women making challenging arguments during disagreements: claiming that their emotions interfere with the validity of what they say. By attributing a woman’s arguments to emotionality, people assume she is unable to think clearly or rationally, and as a result, makes weak arguments. In fact, a recent study found that 13% of Americans still believe that men are better suited emotionally for politics than women (Carnevale et al., 2019). Thus, for every eight voters, one may be using biased perceptions of candidates’ perceived emotionality to guide their decisions about who should lead United States (U.S.) politics.

These examples reveal the consequences of wielding stereotypes and labels in highly charged situations. However, the effects of leveraging emotionality labels in everyday arguments have yet to be studied. Therefore, we wanted to understand the effects of someone intentionally using the “emotional” label to delegitimize their conversation partner’s argument in minor-stakes scenarios. Consistent with research on incivility about how seemingly minor biased comments and actions can create substantial negative consequences for women (Schilpzand et al., 2016), we explored how the label of emotional can be used to undermine and invalidate women’s arguments, and in turn how their progress in the workplace is impeded. In a series of studies, we sought to answer the following questions: how does being labeled as emotional affect perceptions of the legitimacy of a woman’s arguments, and what are the downstream consequences of such perceptions?

¹Pennsylvania State University, University Park, PA, USA
²Ramapo College of New Jersey, Mahwah, NJ, USA

Corresponding Author:
Teresa J. Frasca, 140 Moore Building, University Park, PA 16802, USA.
Email: tuf213@psu.edu
We present a brief review of the well-established literature on stereotypes of women as emotional and men as emotion-ally restrained or inexpressive. Then, we turn to a discussion of how the emotional label relies on ambiguous circumstances and the activation of such stereotypes to delegitimize women’s arguments, and we acknowledge important implications for women’s workplace success.

**Gender, Emotion, and Legitimacy**

Women’s inability to properly control emotions is one of the most salient and consistent stereotypes in the West (Durik et al., 2006; Fischer & Manstead, 2000; Shields, 2016). Indeed, emotion has a history of being associated with femininity and “irrational, disorganized behavior” (MacArthur & Shields, 2015, p. 40), with women being seen as dispositionally, and therefore chronically, emotional and irrational (Barrett & Bliss-Moreau, 2009; Brescoll, 2016). Such judgements are formed around beliefs that emotions are antithetical to rationality, as generally, people tend to see the two as mutually exclusive (Averill, 1980; Ben-ze’ev, 2000). An emotional label, regardless of the actual emotionality of a person, then, may increase skepticism of that person’s ability to be rational. Put simply, if someone suggests that a woman is emotional, then they are also suggesting that she cannot be thinking rationally.

The strength and danger of these stereotypes is reflected in social phenomena like gaslighting, where abusers often label their victims as emotional to invalidate accusations of abuse (Sweet, 2019). By saying that their victim is emotional, the abuser calls the rationality of the victim’s argument into question—a strategy used consciously to discredit the victims’ concerns and claims (Sweet, 2019). In leveraging this stereotype-laden label, abusers can continue their abuse with any observers writing off the victim’s claims as irrational. As such, calling women emotional functions to delegitimize women’s arguments, that is, making an argument seem invalid from the perspective of an outside observer (Berger et al., 1998; Zelditch, 2000). We argue that the emotional label can be used similarly even in non-violent contexts, such as workplace disagreements, to delegitimize a woman’s claims about her conversation partner’s actions.

We do not expect an emotional label to activate the same concepts and supposed irrationality when used against men. Although men in Western cultures are stereotyped as inexpressive (MacArthur & Shields, 2015) and criticized for being unemotional (see Shields, 2013 for a review), under some circumstances men’s emotional expression is allowed and even applauded compared to women’s. Men who are overcome with restrained emotion—meaning maintaining a “stiff upper lip” in certain contexts where they are seen as trying to withhold emotional expressions—are judged less severely than women in the same situations (see MacArthur & Shields, 2019 for a review). Men get the benefit of the doubt with this “passionate restraint.” Moreover, men who cry in masculinized sporting contexts and occupations (e.g., weightlifting or firefighting) are seen as more emotionally strong and their tears are deemed more emotionally appropriate than men who cry in feminized contexts and occupations (e.g., figure skating or nursing; MacArthur, 2019). In a disagreement where an emotional label is employed against a man, his argument may not be delegitimized to the same extent as a woman’s because his emotions are not seen as uncontrollable as hers—the same stereotype does not exist.

In line with these findings, prior work has found that participants associate men with more emotional competence than women (Hess et al., 2016; Shields & Crowley, 1996). For example, Timmers et al. (2003) found that participants believed men are better than women in situations where sensitivity is required for competence, such as in nursing, although one study did find that men who were described as working in a job that required emotional skill and sensitivity were seen as more insecure and wishy-washy than a woman in the same job (Heilman & Wallen, 2010). And while past studies have found that participants generally view shedding tears to signal a loss of control (Vingerhoets et al., 2000), participants rate men who cry in modest amounts more positively than women who cry similar amounts (Warner & Shields, 2007).

In addition, research and theorizing on precarious manhood argues that masculinity is defined by three mandates, one of which is *anti-femininity* (Vandello & Bosson, 2013). Thus, when men act in stereotypically feminine ways (e.g., a way that is seen as overly emotional), they will incur backlash from others for violating masculinity doctrines. Similarly, another study found that participants strongly believed women should express emotion and that men are strongly *proscribed* from being emotional (Prentice & Carranza, 2002). In this way, being labeled as emotional should activate stronger stereotypes about women, compared to men, who may get the benefit of the doubt due to such proscriptions and anti-femininity mandates.

Finally, direct comparisons of men and women’s emotional expression at work supports our prediction that men labeled as emotional will get the benefit of the doubt compared to women labeled similarly. Brescoll and Uhlmann (2008) found that participants perceived women leaders who expressed anger to be more out of control than men leaders who expressed the same levels of anger. This belief that men emote more competently than women for the same emotions leads to several consequences, such as women being perceived as less successful in the workplace (Fischbach et al., 2015), women being granted less power and status, and justifying women’s lower pay (Brescoll & Uhlmann, 2008).
Prior theorizing and empirical work suggest that significant consequences may depend on perceptions of one’s legitimacy in the workplace. For example, Vial et al.’s (2016) theoretical model on the self-reinforcing cycle of illegitimacy posits that if a leader is seen as illegitimate (and in particular, a woman leader), then the leader will experience backlash and insubordination from followers. Moreover, the model suggests that those insubordinates will be less likely to go above and beyond their job demands with behaviors that support their leader (e.g., working late). Empirical work has also shown that perceptions of legitimacy increase compliance with requests from and deference to authorities, and acceptance of information (Johnson et al., 2006; Levi et al., 2009). However, this body of literature examines insubordination when a woman is already an established leader. Our current studies extend such work by investigating how being labeled as emotional may affect her perceived legitimacy and advancement opportunities rather than experiences as a current leader.

To this end, prior research has studied the delegitimizing function of an emotional label in heightened situations such as gaslighting, where the consequences of delegitimization are often physically dangerous and pressing. However, we are interested in how seemingly minor incidences of using the label affect women’s everyday experiences in the workplace. Everyday emotion may be more ambiguous, complicating judgments of emotionality and rationality. Shields (2005) argues that emotions are visible but also ambiguous, fleeting, and complex in everyday interactions. People can mask some emotions, exaggerate others, and express multiple emotions at once. This dynamic, temporary nature of emotions makes it difficult for assertions regarding its quality, quantity, and appropriateness to be countered with objective evidence (Gross et al., 2000; Shields, 2005), which contributes to women’s difficulty in combatting emotional labels once they are employed. Indeed, it is in ambiguous situations, rather than more clear-cut situations, that unconscious biases come out in social judgments (e.g., Barrantes & Eaton, 2018).

Building on this ambiguity is how an emotional label increases the salience of gender-based emotion stereotypes. Ambiguous situations provide ample opportunity for stereotypes to be unconsciously activated and to influence judgments of when emotions are appropriate and inappropriate. For example, Vial et al. (2016) argue that increased gender salience activates negative stereotypes about women and hurts them as leaders. Pulling from research on stereotype threat, even subtle environmental cues to stereotypes have been shown to create inequalities (e.g., Seitchik et al., 2014; Spencer et al., 1999). A subtle nod to stereotypes about women’s overemotionality should, in theory, create the same effects—and in the scenarios we are interested in, explicitly labeling a woman as emotional is more than a subtle, “in the air” nod to these stereotypes. In sum, we argue that the uncertainty of emotions, their vulnerability to biases, and the ability of a label—and in particular, an emotional label—to highly increase the salience of overemotionality stereotypes about women create a particularly fertile breeding ground for the delegitimization of women’s arguments even in more minor everyday interactions in the workplace.

In conclusion, prior work has demonstrated robust stereotypes about solely women’s overemotionality and thus irrationality, making women’s arguments especially vulnerable to being delegitimized when labeled as emotional, regardless of their actual emotionality. The current research is unique in that instead of examining the effects of the actual emotionality of men and women targets, as prior work has done, we tested how the simple label of emotionality may create differences in how we view men and women during disagreements. Put simply, we were interested in how a label that is wielded like a weapon during an argument can hurt perceptions of women’s legitimacy. We sought to answer two questions as yet untested in the literature: are women actually delegitimized more than men when a conversation partner introduces an emotional label during a disagreement, regardless of their actual emotionality? And if so, what are the consequences?

**Overview of Present Studies**

In the present research, we examined observers’ perceptions of the legitimacy of a woman’s argument after she is called emotional during a disagreement. The combined literature on gender, legitimacy, and emotion suggests that when observers view an exchange where a woman is labeled emotional, they will subsequently perceive her argument to be less valid than a similarly labeled man’s argument. In addition, we predicted that, when women (versus men) are labeled emotional, perceived emotionality would mediate the gender-legitimacy relationship. Specifically, gender coupled with an emotional label cues the outside observer’s judgment of women’s emotionality by activating the stereotype that women are overemotional and irrational whereas men are not stereotyped similarly. Therefore, we predicted that women’s arguments would be delegitimized when called emotional more than when the label is absent. To rule out the possibility that women are perceived as emotional irrespective of the emotional label, we included a control condition where women and men were not called emotional. To that end, we expected no differences in perceived legitimacy between men and women’s arguments when no emotion label was provided.

When calling someone emotional, delegitimization occurs through a process involving three contributors: the **labeler**—the person who calls the other emotional, the **target**—the one being called emotional, and the **observers**—the individuals who determine, or confer, legitimacy. In the studies that follow, participants serve as the observers.

**Study 1**

In Study 1, we used an experimental vignette methodology to test three hypotheses. The vignette involved two individuals...
in a disagreement, culminating with the labeler either calling the target emotional, or simply continuing the disagreement. We hypothesized:

**H1:** When the target is labeled as emotional by a conversation partner, participants would perceive the woman’s argument as less legitimate than the man’s argument, whereas no differences would emerge between women and men targets in the control condition where no emotion label is given.

**H2:** Participants would perceive the women’s argument as less legitimate when she is labeled emotional than in the control condition.

**H3:** Participants’ perceptions of the target’s emotionality would emerge as a mediator between condition and the participant’s perception of the argument’s legitimacy.

In exploratory analyses, we examined the relation between participant gender and their perceptions of the legitimacy of the labeler’s arguments to explore the possibility that, when women targets are called emotional, women participants would not perceive their arguments as less legitimate than women targets’ arguments in the control condition. Previous research has found gender differences in third-party perceptions of microaggressions toward women in the workplace—namely, that women participants detected greater discrimination than men when reading vignettes of subtle, potentially discriminatory male supervisor behavior toward female subordinates (Basford et al., 2013). Thus, when viewing someone calling a woman emotional, women participants may detect this subtle bias and subsequently refrain from delegitimating her argument. However, other research finds that individuals, irrespective of gender, perceive women according to gender-based stereotypes (Fischer & Manstead, 2000). Thus, if women and men participants both perceive women targets as emotional when labeled as emotional, participant gender differences in our predicted effects may not occur. We included participant gender as an exploratory variable to see if these gender differences manifested.

**Method**

**Participants**

Eighty-six undergraduate students at a small, Mid-Atlantic university participated for course credit. Two participants were excluded from the analyses for failing to attend to the questions, defined as ignoring reverse-scored items throughout the questionnaire. The final sample consisted of 48 women (57.1%) and 36 men (42.9%). Most participants self-identified as White (n = 69, 82.1%), followed by Mixed (n = 6, 7.1%), Asian American (n = 4, 4.8%), African American (n = 3, 3.6%), and Latina/o (n = 2, 2.4%).

**Procedure**

Stimuli and measures for all three studies can be found in online Supplemental Materials. Participants read a vignette about two individuals having a disagreement over a group project. In the experimental condition, the vignette ends with the labeler implying the target is emotional during the disagreement. In the control condition, the labeler simply disagrees with the target with no emotion label (see script below). We manipulated target gender by using names informally pretested to be associated with men (Tom) or women (Kim); however, the labeler gender was always a man (Jason). Participants were instructed to think about the situation as if it were really happening and to provide feedback as if they were outside observers to this situation. The vignette is as follows, with the **bolded phrase** representing the experimental condition, and the italicized phrase representing the control:

Kim/Tom and Jason are working on a group project. They are arguing about contributing to the project equally.

**Kim/Tom:** This is difficult to say but I’m beginning to think you aren’t pulling your weight in the project.

**Jason:** What do you mean?

**Kim/Tom:** It’s just that I’ve been doing a lot of research and it seems like you’ve barely started your part of the project.

**Jason:** Well, if you feel that way then you haven’t noticed the work I’ve been doing.

**Kim/Tom:** But I should be seeing the work by now. The project is due soon!

**Jason:** But we didn’t decide that today is the day for us to put everything together. If you would have told me that you wanted to see my stuff today, I would have brought it.

**Kim/Tom:** That just doesn’t seem like a good enough excuse for me.

**Jason:** **Hey - why don’t you just take a moment to calm down?** OR **Hey—why don’t you see my point of view on this?**

**Measures**

After reading the vignette, participants responded to a series of questions. To assess participants’ perceptions of the validity of the target’s arguments, we created an 8-item measure of perceived validity. Sample items included, “[Kim’s/Tom’s] opinion was valid” and “[Kim’s/Tom’s] opinion makes sense” (Cronbach’s α = 0.89). Items were rated on a
7-point scale from 1 (strongly disagree) to 7 (strongly agree). The same 8-item measure of perceived validity was used to examine the perceived legitimacy of Jason, the labeler’s arguments (Cronbach’s α = 0.90). Question order was counterbalanced in terms of whether the participants rated the validity of the target’s or labeler’s arguments first. Finally, participants rated the target’s perceived emotionality using a single item (How emotional is [Kim/Tom]?) on a 7-point scale from 1 (not at all) to 7 (very much). This item was embedded within a matrix we created containing 10 other emotions that we did not analyze (e.g., annoyed, joyful, nervous) to dispel participant suspicion about the purpose of the study.

Results

To test Hypotheses 1 (H1) and 2 (H2), we conducted a 2 × 2 (Target Gender: Man or Woman) × 2 (Labeler’s Evaluation: Emotional or Control Condition) between-subjects analysis of variance (ANOVA) with the perceived validity of the target’s arguments as the dependent variable. Participant gender was entered as an exploratory subject variable. This analysis yielded a significant main effect for participant gender such that women participants perceived the targets’ arguments as more legitimate than men participants did (M = 4.75, SD = 0.82 vs. M = 4.30, SD = 1.25, respectively), F(1, 76) 4.32, p = .04, ηp² = 0.05. No other effects involving participant gender emerged.

Consistent with H1, the interaction between labeler’s evaluation and target gender was significant, F(1, 76) 4.63, p = .04, ηp² = 0.06 (see Figure 1). When the labeler evaluated the woman target as emotional, participants perceived her arguments to be less legitimate relative to when the labeler evaluated the man target as emotional, (M = 4.01, SD = 1.19 vs. M = 4.82, SD = 1.05, respectively), F(1, 76) 7.16, p = .009, ηp² = 0.09. No differences emerged between the men and women targets in the control condition (M = 4.68, SD = 0.97 vs. M = 4.66, SD = 0.87, respectively), F(1, 76) 0.12, p = .73, ηp² = 0.002. Also in support of H2, participants viewed the woman target’s arguments as less legitimate in the emotional condition relative to the control condition, F(1, 76) 6.22, p = .02, ηp² = 0.08.

To test H3 and examine whether perceptions of the target’s emotionality mediated the relation between the labeler’s evaluation of the target and the participant’s perception of the legitimacy of the target’s arguments, we conducted a mediation analysis using Preacher and Hayes’ (2008) PROCESS Macro. Using 5000 bootstrap resamples with a 95% confidence interval (CI), which is recommended for small samples, we compared the participants’ perceptions of the woman target who was called emotional against the

---

**Figure 1.** Mean Perceptions of Target’s Argument Legitimacy by Condition, Across Studies 1–3 (Left to Right).

---

Note. †p = .06, *p < .05, **p < .01, ***p < .001.
other conditions using PROCESS Model 4. The labeler’s evaluation and target gender were set up as a contrast (−3 = women targets called emotional, 1 = men targets called emotional, 1 = men targets in the control condition, and 1 = women targets in the control condition). This variable was entered as the predictor (X), target’s perceived emotionality was entered as a mediator (M), and the perceived legitimacy of the target’s arguments was entered as the outcome variable (Y; see Figure 2).

Consistent with H3, the contrast variable significantly predicted the target’s perceived emotionality, and the target’s perceived emotionality in turn predicted the perceived legitimacy of the target’s arguments. The 95% bias-corrected CIs [−0.35, −0.005] for the target’s perceived emotionality did not include zero, indicating that the target’s perceived emotionality mediated the relation between women targets being called emotional and how legitimate her arguments are perceived. In other words, when the labeler called the woman target emotional, participants perceived the women targets to be more emotional than all other conditions, and this perceived emotionality, in turn, led participants to perceive the targets’ arguments as less legitimate than all other conditions.

**Exploratory Analysis**

Scholarship on legitimacy suggests that the legitimacy of both individuals in an argument is contingent on the perceptions of the observers who view their argument (Berger et al., 1998). To address this facet of our scenarios, we conducted an exploratory analysis to assess whether labeling the target as emotional affected participants’ perceptions of the legitimacy of the labeler’s arguments.

We conducted a 2 (Target Gender: Man or Woman) x 2 (Labeler’s Evaluation: Emotional or Control Condition) x 2 (Participant Gender: Man or Woman) between-subjects ANOVA with the perceived validity of the labeler’s arguments as the dependent variable. There were no significant main effects, but the interaction between labeler’s evaluation and target gender emerged as significant, F(1, 76) 4.74, p = .03, η²p = 0.06. Participants rated labelers as marginally more legitimate when they evaluated women targets as emotional than when evaluating men targets as emotional, (M = 4.84, SD = 1.07 vs. M = 4.21, SD = 1.21), F(1, 76) 3.40, p = .07, η²p = 0.04. While the means did not reach significance at the p < .05 level, their direction reflects that participants also rated labelers’ arguments as marginally more legitimate in the man target control condition than when they evaluated men as emotional, (M = 4.77, SD = 0.82 vs. M = 4.21, SD = 1.21), p = .10, η²p = 0.03. No differences emerged when comparing women targets in the emotional versus control conditions (M = 4.84, SD = 1.07 vs. M = 4.40, SD = 1.09), F(1, 76) 2.08, p = .15, η²p = 0.03.

**Discussion**

Results from Study 1 supported our hypotheses. Consistent with H1 and H2, when the labeler evaluated the woman target as emotional, participants perceived the woman target as making less legitimate arguments than women in the control condition and men targets labeled as emotional. Furthermore, supporting H3, women labeled emotional were perceived to make less legitimate arguments because participants perceived them to be more emotional than men in the experimental and control conditions and women in the control condition. This study provides evidence that when someone labels a woman as emotional, those witnessing this label may use this information as they assign value to her arguments. However, this very same evaluation does not appear to have the same consequences for men. Notably, while we observed a main effect of participant gender, women participants did not differ from men participants in
their perception of women targets’ arguments when they were called emotional. These findings are consistent with previous research demonstrating that both men and women hold stereotypic views of women and men’s emotionality (Fischer & Manstead, 2000) and inconsistent with the prospect that women would be more likely to detect subtle bias than men (Basford et al., 2013).

While these results support our predictions, the study has some limitations. First, we presented participants with one vignette, creating the possibility that the results emerged due to vignette idiosyncrasies. Also, in this study, the labeler is always the one on the defensive and the target is always on the offensive (i.e., in all conditions, Jason is defending against Tom/Kim’s argument that he failed to prepare adequality for their project due date). In addition, our exploratory analyses on the effects of labels on the labeler’s perceived legitimacy warrant further investigation. Compared to control conditions, participants perceived the labeler’s argument to be most legitimate when the labeler called the woman target emotional and least legitimate when the labeler called the man target emotional. Study 2 will help us disentangle if this finding is due to a contrast effect (where labeling a target as emotional also boosts the perceived legitimacy of the labeler) or the result of the emotional label seeming unrealistic for men targets. Finally, we wanted to explore whether labeler gender affects the perceived legitimacy of their arguments—namely, if it matters that a man versus a woman labels the target as emotional.

**Study 2**

To address limitations in the first study, as well as extend our analysis of the perceived legitimacy of the labeler’s arguments, we made several changes to the vignette for Study 2. First, to control for the possibility that idiosyncrasies of the vignette topic may have influenced the results, we changed the vignette topic. Second, we counterbalanced the script by varying the first character to speak (e.g., the labeler or the target), which unconfounded argument content with labeler’s evaluation. Third, we varied the gender of the labeler. Finally, rather than telling the target to “calm down,” the labeler explicitly describes the target as emotional. This modification of the language for Study 2 allowed us to examine more directly the effect of being labeled as emotional. The term emotional, in many mainstream American contexts, is associated with uncontrolled actions and irrationality. For example, Shields and Crowley (1996) found that when a target’s responses were described as emotional they were judged as less controlled and less appropriate than when those responses were described by specific emotional terms (e.g., angry).

For this study, we maintained our three hypotheses from Study 1 and added one more to incorporate the results from Study 1’s exploratory analyses:

**H4:** Participants’ perceptions of the legitimacy of the labeler’s arguments would emerge as a mediator, such that the labeler’s evaluation of the woman target as emotional would lead her arguments to be delegitimized relative to all other conditions, because participants would perceive the labeler’s arguments as more legitimate than in all other conditions.

**Method**

**Participants**

One hundred twenty-six undergraduate students at a small, Mid-Atlantic university participated for course credit. Consistent with Study 1, we removed two participants who failed to attend to the questions, defined as ignoring reverse-scored items throughout the questionnaire. The final sample consisted of 60 women (48.4%) and 64 men (51.6%). Most participants self-identified as White (n = 106, 85.5%), followed by African American (n = 6, 4.8%), Asian American (n = 5, 4.0%), Latina/o (n = 5, 4.0%), Native American (n = 1, 0.8%), and Mixed (n = 1, 0.8%).

**Procedure and Materials**

The procedure and materials remained the same as in Study 1 except for a few changes. The vignette content involved a school club budget (as opposed to a group project), and we varied how the labeler calls the target emotional (i.e., “You’re being emotional about this” instead of Study 1’s “Why don’t you just take a moment to calm down?”). In addition, target genders were associated with new names and labeler gender was varied across participants, resulting in a 2 (Target Gender: Man or Woman) x 2 (Labeler Gender: Man or Woman) x 2 (Labeler’s Evaluation: Emotional or Control Condition) x 2 (Participant Gender: Man or Woman) between-subjects design. The vignette is

| Table 1. Examples of Counterbalancing the Opening Lines in Study 2’s Vignettes. |
|---|---|
| **Example 1** | **Example 2** |
| John says, “I cannot afford this. You handled the money and should have warned us that we were running over budget.” | Amy says, “I cannot afford this. You handled the money and should have warned us that we were running over budget.” |
| Amy says, “There was just no way that I could have known. Those costs happened at the last minute.” | John says, “There was just no way that I could have known. Those costs happened at the last minute.” |
| John then says, “You’re being emotional about this.” | John then says, “You’re being emotional about this.” |

Note. Names could also be Lisa or Matt, as described in the text above.
as follows, with the **bolded phrase** representing the experimental condition, and the *italicized phrase* representing the control:

John/Lisa and Matt/Amy are student officers for a martial arts club on campus. They planned an intercollegiate competition where they had to raise money to pay for judges, materials, and space. Now that the competition was over, they realize that they had spent $350 over their budget, which means that they have to pay $175 each out of their own pockets.

John/Lisa says, “I cannot afford this. You handled the money and should have warned us that we were running over budget.”

Matt/Amy says, “There was just no way that I could have known. Those costs happened at the last minute.”

John/Lisa/Matt/Amy then says, “**You’re being emotional about this.**” OR “**I just don’t agree with you.**”

We counterbalanced the vignettes such that for half of the participants, the labeler stated the first line, and for the other half, the target stated the first line (meaning the other character then said the second statement). The labeler always stated the last line (see Table 1 for clarifying examples, as well as the online Supplemental Materials). In this way, we varied whether it was the labeler or the target who was making the defensive argument. As with Study 1, we also counterbalanced the order in which participants rated the legitimacy of the labeler’s or target’s arguments.

We used the same measures of perceived legitimacy of the target’s arguments (Cronbach’s α = 0.91) and perceived legitimacy of the labeler’s argument (Cronbach’s α = 0.90) as Study 1, including the emotionality item (again embedded in a matrix of items) to assess the target’s perceived emotionality.

**Results**

Our exploratory analyses revealed no significant differences between labeler genders, or participant genders for key variables, all ps > .05. Thus, results are combined across labeler and participant gender. We included whether the participant was asked first about the legitimacy of the target’s arguments or the legitimacy of the labeler’s arguments as a covariate.

**Hypotheses 1 and 2**

To test H1 and H2, we conducted a 2 (Target Gender: Man or Woman) x 2 (Labeler’s Evaluation: Emotional or Control Condition) between-subjects ANOVA with the perceived legitimacy of the target’s arguments as the dependent variable. Consistent with H1 and replicating Study 1, a significant interaction emerged, \(F(1,116) = 8.78, p = .004, \eta^2_p = 0.07\) (see Figure 1). When the labeler called the woman target emotional, participants assessed her arguments as less legitimate relative to when the labeler evaluated the man target as emotional, \((M = 4.16, SD = 1.44\) vs. \(M = 5.13, SD = 1.05\), respectively), \(F(1,116) = 8.64, p = .004, \eta^2_p = 0.07\). Also consistent with H1, no differences emerged between the men and women targets in the control condition \((M = 4.79, SD = 1.27\) vs. \(M = 4.40, SD = 1.37\), respectively), \(F(1,116) = 1.54, p = .22, \eta^2_p = 0.01\). In support of H2 and replicating Study 1, when the labeler evaluated the woman target as emotional, her arguments were seen as marginally less legitimate relative to the woman target in the control condition, \(F(1,116) = 3.51, p = .06, \eta^2_p = 0.03\). Finally, when the labeler evaluated the man target as emotional, his arguments were seen as more legitimate than the control condition, \(F(1,116) = 5.37, p = .03, \eta^2_p = 0.04\).

**Hypotheses 3 and 4**

To test H3 and H4, we conducted a mediation analysis using Preacher and Hayes’ (2008) PROCESS macro (Model 4), setting 95% CIs and using 5000 re-samples. Labeler’s evaluation and target gender were set up as a contrast comparing woman targets called emotional with all others \((-3 = \text{women targets called emotional, } 1 = \text{men targets called emotional, } 1 = \text{men targets in the control condition, and } 1 = \text{women targets in the control condition})\). We entered this variable as the predictor (X). We entered the participants’ perception of the target’s emotionality (M1) and legitimacy of labeler’s arguments (M2) both as mediators, and the legitimacy of the target’s arguments as the outcome variable (Y; see Figure 3).

We included, as a covariate, whether or not the first or second character to speak in the dialog was the labeler, and thus the one to say “you’re being emotional.”

The contrast variable significantly predicted the target’s perceived emotionality, \(b = -0.17, t = -2.27, p = .03\), but did not significantly directly predict legitimacy of the labeler’s arguments, \(b = -0.11, t = -1.84, p = .07\), although means were in the same direction as Study 1. The participants’ perception of the target’s emotionality and the legitimacy of the labeler’s arguments did significantly predict their perception of how legitimate the target’s arguments were, \(b = -0.07, t = -2.72, p = .007\) and \(b = -0.50, t = -5.77, p < .001\), respectively. Finally, to establish mediation, we looked at the indirect effects for both mediators. The 95% bias corrected CIs for the target’s perceived emotionality did not include zero [0.008, 0.08], indicating that the target’s perceived emotionality mediated the relation between woman targets being called emotional and the legitimacy of the target’s arguments and supporting H3. In addition, the results supported H4 because the 95% bias corrected CIs for the legitimacy of the labeler’s arguments did not include zero [0.002, 0.13], indicating that the legitimacy of the labeler’s arguments mediated the relation between
woman targets being called emotional and the participants’ perceptions of the legitimacy of the target’s arguments.

In sum, when the labeler called the woman target emotional, participants perceived the woman targets’ arguments as less legitimate than all other conditions, because (1) participants perceived the woman in this condition to be the most emotional and (2) because participants perceived the labeler’s comments as more legitimate when calling a woman emotional than when the labeler engaged with the targets in any of the other conditions.

**Discussion**

Consistent with H1–H4 and replicating the results of Study 1, when the labeler called the woman to target emotional, participants perceived the woman’s arguments as less legitimate than when the labeler called the man target emotional and simply disagreed with the men and women targets in the control condition. In Study 2 we also found that the man target’s arguments were seen as more legitimate when the labeler called him emotional than in the control condition. Mediational analyses were consistent with predictions; when the labeler called the woman emotional, participants perceived her as conveying less legitimate arguments because they perceived her as emotional and because the labeler’s arguments were seen as most legitimate. In other words, when someone calls a woman emotional, their own arguments may be seen as more valid. Finally, consistent with Study 1, participant gender did not interact with our predicted pattern of results, providing further evidence that men and women participants view women called emotional similarly to one another. Further, labeler gender did not impact our results.

While Study 2 replicated many of the findings from Study 1, limitations remain. In terms of statistical power, both Studies 1 and 2 involved small samples. In addition, the vignettes in Studies 1 and 2 were relatively short and were written dialogs, allowing participants to inscribe their own interpretation of the situation’s tone onto the vignette. Finally, Studies 1 and 2 demonstrated that being labeled as emotional is particularly harmful for perceptions of women’s argument legitimacy but did not assess potential consequences such a label might have for women. Study 3 addressed each of these limitations by increasing the sample size, changing the written dialog to a video format, and measuring potential consequences.

**Study 3**

Study 3 mirrored Studies 1 and 2 with a few key changes. First, to simplify analyses and because Study 2 did not find any differences based on labeler gender, we returned to Study 1’s design in which the labeler is always a man. Second, we used the group project vignette from Study 1 as a base and expanded the script so that participants could get more information to inform their judgments. Third, we altered the control condition such that we simply ended the script with the target’s last statement (instead of the labeler stating, “I just don’t agree with you”) to test the effects of a label alone, instead of in comparison to a continued contestation by the labeler. Fourth, we created and used a video version of the vignette. As Sleed and colleagues (2002)
argue, in video-vignettes, “all of the contextual information is given, and it is up to the subject to choose their points of focus rather than base their decisions on aspects of the context that researchers deem to be of importance” (p. 23). Videos may be more externally valid than written vignettes in this way. Moreover, previous research has shown videos to be highly engaging and perceived as highly realistic by viewers (Visser et al., 2016). Finally, in one study of victim-blaming in rape scenarios, participants were significantly more likely to blame the victim and less likely to see the scenario as rape when they read a written vignette versus watched a video vignette (Sleed et al., 2002). Thus by using video vignettes in Study 3, we hoped to improve the external validity of our tests, improve engagement, and test if the effect held outside of a written description methodology. A White man played the labeler, and a White woman and man played the two targets (see Method section below for more information and the online Supplemental Materials). Finally, given prior theory and research suggesting that delegitimization has negative consequences for leaders such as decreased compliance, cooperation, and deference from subordinates (e.g., Johnson et al., 2006; Vial et al., 2016) and that being perceived as emotional negatively affects workplace success and compensation for women (e.g., Brescoll & Uhlmann, 2008; Fischbach et al., 2015), we added outcome-based measures to Study 3 to capture the potentially negative effects of being labeled as emotional. In this way, we sought to understand how an emotional label might extend beyond one conversation and affect women’s advancement in other arenas.

The first four hypotheses for Study 3 were the same as those for Study 2 (H1–H4). We also added one more hypothesis:

H5: Women targets labeled as emotional would face more negative groupwork outcomes resulting from their arguments being perceived as less legitimate.

Method

Participants

To address the lack of power in Studies 1 and 2, we conducted an a priori power analysis in G*Power using effect sizes from the prior studies to determine how many participants we needed to recruit for Study 3. Analyses in G*Power (Faul et al., 2009) estimating an effect size of 0.22 and an alpha error probability of 0.05 revealed that 212 participants would be needed to achieve a power of 0.8 with this design. Based on previous studies run with this participant pool, we expected about 20% of our sample to provide incomplete data or fail the attention check, so we collected data from an additional 40 participants. Two hundred fifty-one undergraduate students at a large, Mid-Atlantic university participated for course credit. Consistent with Studies 1 and 2, prior to data analyses we removed participants who failed both attention checks (n = 16), failed to correctly recall key information from the scene (n = 18), and/or completed the survey multiple times (n = 2), resulting in a final sample size of 218 (M_{age} = 18.75, SD = 1.23). Seventy-one percent of participants identified as women (n = 153), with men (n = 64) constituting 29% of the sample. Most participants self-identified as White (n = 164, 75.2%), followed by Biracial/Multiracial (n = 20, 9.2%), Asian American (n = 16, 7.3%), Hispanic/Latinx (n = 11, 5.0%), and African American (n = 6, 2.8%).

Procedure and Materials

The procedure and materials were similar to Study 1. We used a video version of the vignette from Study 1, where participants viewed a scenario in which two students (the labeler, Jacob; and a target, Abby/Dylan) disagreed over a group project (see online Supplemental Materials). Character names were pretested as representing men and women aged 18–24 years (Newman et al., 2018). In the experimental condition, the vignette ends with the labeler calling the target emotional. The control condition video is the exact same video, just with the labeler’s final lines (including the emotional label) cut off. Participants could watch the video as many times as they needed, and they also received a transcript of the dialog:

Abby/Dylan: Okay, our final project is due in three days. I finished my slides and my half of the paper yesterday, did you get a chance to look at them?
Jacob: Yeah it looks good to me. There are just a few things you might want to change once we have the whole paper and presentation together.
Abby/Dylan: What about your slides?
Jacob: I have a bunch of stuff this week, I’m going to finish them soon.
Abby/Dylan: And your part of the paper?
Jacob: I have it outlined, and it won’t take me long to write it all out.
Abby/Dylan: Ughh… you gotta start pulling your weight in the project.
Jacob: What do you mean?
Abby/Dylan: It’s just that I’ve been doing a lot of research and it seems like you’ve barely started your part of the project.
Jacob: Well, if you feel that way then you haven’t noticed the work I’ve been doing.
Abby/Dylan: But I should be seeing the work by now. The project is due soon!

Jacob: But we didn’t decide that today is the day for us to put everything together. If you would have told me that you wanted to see my stuff today, I would have brought it.

Abby/Dylan: That just doesn’t seem like a good enough excuse for me. This project is worth 30% of our grades, and I really want to do well on it.

Jacob: I want to do well on it, too. I’ve just been busy. We still have 3 days.

Abby/Dylan: I know, but I’ve also been busy and I’ve still found time to complete my portion with enough time for us to both go over it together before turning it in. I don’t think it’s fair that I did all of my work and it sounds like you’ve barely started your half.

(Experimental Condition Only)

Jacob: Hey—why don’t you just take a moment to calm down? You’re getting emotional!

As with Studies 1 and 2, participants were instructed to think about the situation as if it were really happening and to provide feedback as if they were outside observers of this situation. Videos were pilot tested using a within-subjects design with 20 participants from the same sampling pool, revealing no significant differences in how many times participants watched the video, how likely, vivid, and realistic the scenario was, and how believable each character was, all ps > .09. Participants did not perceive the actors to be of significantly different ages, p = .09, and all participants accurately perceived Abby to be a woman and Dylan and Jacob to be men.

Measures

We counterbalanced all measures in terms of whether the participants provided ratings for the target or the labeler first. Measures of perceived legitimacy of the target’s argument (Cronbach’s α = 0.93) and perceived legitimacy of the labeler’s argument (Cronbach’s α = 0.93) were identical to those used in Studies 1 and 2. Study 3 again used the same single-item measure of emotionality embedded within the matrix of 11 emotions.

To extend the first two studies and assess how the perception of emotionality as illegitimacy may create downstream consequences for women labeled as emotional, we also assessed how being labeled may affect groupwork outcomes. Importantly, we designed these items such that they were relevant to the context of the vignette, relatable to our student population, and at least somewhat analogous to more traditional measures of workplace outcomes.

First, participants assigned each character a grade between 0 to 100 points. Next, using a 7-point Likert scale ranging from 1 (not at all) to 7 (very), participants rated how likely they would be to choose each character to be their group leader. Participants then answered a yes or no question asking whether or not they would blame each character if the group were to receive a bad grade and the extent to which they would blame them on a scale from 1 (not at all blameworthy) to 7 (extremely blameworthy). Participants also rated how likely two specific outcomes would be in future interactions with the target as the team lead (“Your ideas would be dismissed,” and “Your ideas would be respected” [reverse-coded], r = .56, p < .001). Finally, participants responded to three items assessing their willingness to go above and beyond if they were the target’s subordinate (reflecting prior work cited in our introduction on how women leaders receive less voluntary, extra-role support from subordinates), including, “You would spend extra time working on the project if needed,” and “Other members of the group would undermine Abby/Dylan’s leadership.” All leadership items were rated on a 7-point scale from 1 (not at all likely) to 7 (extremely likely).

Results

We found no significant effect of condition on work outcomes, and therefore no support for H5 regarding the downstream consequences of labeling. Specifics of all H5 analyses are reported in detail at the conclusion of the manuscript.²

Hypotheses 1 and 2

Results partially replicate Studies 1 and 2 (see Figure 1). Using the same analyses as Study 2, to test H1 and H2 we conducted a 2 (Target Gender: Man or Woman) x 2 (Labeler’s Evaluation: Emotional or Control Condition) x 2 (Participant Gender: Man or Woman) between-subjects ANOVA with the perceived legitimacy of the target’s arguments as the dependent variable. We found partial support for H1: while there was no significant interaction, p = .32, and no effect of labeler evaluation, p = .58, there was a main effect of gender such that target women were perceived to be significantly less legitimate (M = 3.99, SD = 1.10) than target men (M = 5.43, SD = 1.11), F(1, 214) = 90.61, p < .001, ηp² = 0.30. Analyses of the effect of condition on perceptions of the labeler’s argument legitimacy reflected a similar pattern. No significant interaction emerged, nor did we find a main effect of label on participant’s perceptions that the labeler’s argument was legitimate, all ps > .11; however, we found a main effect of gender such that labelers’ arguments were perceived as more legitimate when the target
was a woman \( (M = 5.24, SD = 1.01) \) than when the target was a man \( (M = 3.85, SD = 1.24) \), \( F(1, 214) = 82.94 \), \( p < .001 \), \( \eta^2_p = 0.28 \).

**Hypotheses 3 and 4**

As with Study 2, we used Preacher and Hayes’ (2008) PROCESS macro (Model 4) to test H3 and H4. To do so, we entered our condition variable as the predictor \( (X; -3 = \text{women targets called emotional}; 1 = \text{men targets called emotional}; 1 = \text{men targets in the control condition}, \text{and} 1 = \text{women targets in the control condition}) \), the participants’ perception of the target’s emotionality (M1) and legitimacy of the labeler’s arguments (M2) as mediators, and the legitimacy of the target’s arguments as the outcome variable \( (Y) \). Contrary to the results of Study 2, the contrast variable did not significantly predict the target’s perceived emotionality, \( p = .89 \), but did significantly predict the perceived legitimacy of the labeler’s arguments, \( b = -0.14, t = -2.76, p = .006 \). Participant perceptions of the target’s emotionality did not predict perceived legitimacy of the target’s argument, \( p = .20 \), conflicting with analyses replicated in Studies 1 and 2. However, the participant’s perception of the labeler’s legitimacy and the contrast variable did predict perceived legitimacy of the target’s argument, \( b = -0.05, t = -9.51, p < .001 \) and \( b = 0.11, t = 2.63, p = .009 \), respectively, replicating Study 2’s findings. In sum, the target’s perceived emotionality did not mediate the relation between condition and the legitimacy of the target’s arguments (95% bias-corrected CIs for the target’s perceived emotionality \([-0.008, 0.001]\)). By contrast, the legitimacy of the labeler’s arguments did mediate the relation between woman targets being called emotional and the participants’ perceptions of the legitimacy of the target’s arguments (95% bias-corrected CIs for the legitimacy of the labeler’s arguments \([0.028, 0.122]\)).

**Discussion**

Across the novel outcomes that Study 3 incorporates, being labeled as emotional did not have the same consequences the prior literature suggests would be observed and that we predicted would follow. Specifically, in these school-based scenarios, women labeled as emotional did not experience the same judgments of their ability to lead a group project team as would be expected based on Studies 1 and 2 and on prior work on the role of legitimacy in women leaders’ workplace success.

In the analyses in which we did find significant differences, it was most often due to differences in the target’s gender, rather than their gender coupled with the emotional label. The relatively large effect size for the main effect of target gender in Study 3 potentially prevented a detectable interaction between target gender and labeler evaluation. Put more simply, our women targets as a whole may have been perceived as having less legitimate arguments than men targets to such an extent that it did not matter whether or not they were labeled as emotional. Such a large main effect reflects society’s broader tendencies to evaluate women leaders more negatively in comparison to men leaders (e.g., Eagly et al., 1992; Vial et al., 2016). However, there are several limitations to Study 3 that could explain the contradictions between Study 3 and Studies 1 and 2, and that provide an avenue for future research.

The primary difference between the first two studies and Study 3 is the introduction of video-based stimuli. While written vignettes allow us to control for other confounds, video vignettes may assist in painting a more detailed mental picture for participants. Thus, their perceptions of the scenario and the characters within it may have been different with the additional information a video provides. For example, our videos were intentionally created to make participants feel like they were witnessing the disagreement in person, whereas reading Studies 1 and 2’s written vignettes may have made participants feel more like they were observing the interaction over email, a group chat, or other written means. The latter approach introduces more ambiguity—participants are unsure of each character’s body language, their tone of voice, and how they are dressed, for example, all of which may influence the attributions participants make regarding the character. Such a lack of information may cause participants to rely on assumptions and stereotypes and creates space for biases to influence judgements. In the video vignettes however, the witness (i.e., the participant) may not have to rely as much on stereotypes, assumptions, or another’s interpretation of the event to inform their perceptions of the characters, as they have more direct information about each character’s behavior. For example, in the written version of the vignette, a participant may have imagined Abby as sitting with her arms crossed or rolling her eyes while listening to Jacob with an angry facial expression. By contrast, in the video version of the vignette, the participant sees Abby’s facial expressions directly, which may influence their judgment of how emotional Abby is being. In sum, perhaps the actors’ depiction of the scenario in Study 3 was less dramatic than what participants imagined from the written dialog in Studies 1 and 2. In this way, a major reason we did not observe the same pattern of results in Study 3 as we did in Studies 1 and 2 may be the relative differences in contextual information that a video versus a written vignette provides, reflecting potential different aspects and/or boundary conditions of the delegitimization process.

We believe these boundary conditions reflect differing contexts in everyday scenarios as well as specifically in the workplace. Our video vignettes simulated the position of an employee witnessing a disagreement between coworkers. By contrast, our written vignettes may have simulated water cooler or casual conversations that happen at work—such as when that same witness instant messages a coworker who was not present for the disagreement and tells them what was said. Thus, the witness has more details and information
about the situation to influence their judgements than the coworker they instant message. The difference in Study 1 and 2 versus Study 3 findings may therefore indicate that delegitimization occurs more often when a scenario is ambiguous, or when it is only described for someone else, rather than directly witnessed. We speak more to the implications of this potential explanation in our general discussion.

In addition to the different delivery methods corresponding with differences in the observed pattern of results, aspects of the videos themselves may have affected our findings. Although video stimuli were not significantly different from one another in pilot testing, there may have been idiosyncrasies that influenced participants’ ratings. For example, while participants correctly identified the women and men in the videos as their intended gender, it is possible that the actors and actress were not seen as prototypical members of their gender groups. If Abby is seen as a less stereotypical woman, or Dylan and Jacob are seen as less stereotypical men, then perhaps expectations for emotionality and rationality may be less stringent than with more prototypical characters. With differing expectations, an emotional label might not have had as strong a negative impact as it did in Study 1. We did not measure how masculine and feminine participants perceived each character to be, only what they perceived the characters’ gender identities to be, and therefore cannot speak to how these facets may have unintentionally affected our findings.

Finally, another potential idiosyncracy affecting participant judgments could have been the details of the assignment. In Study 1, the target character is upset because Jacob hasn’t completed his portion of the project yet and it is due “soon;” by contrast, in Study 3, the project is due in 3 days. It is possible that, to our undergraduate participants, 3 days is still plenty of time to complete a project, making Abby/Dylan’s frustration with Jacob unreasonable, and making Abby/Dylan equally irrational regardless of an emotional label. In addition, in the first few lines of Study 3’s dialog, Jacob expresses that he has looked at Abby/Dylan’s part of the project, and that he has had a busy week. Such information was not present in the Study 1 dialog. These sentences may have made Jacob’s arguments seem more understandable in Study 3, reducing the impact of an emotional label. Such minor factors may have unintentionally caused differences between the written versus video vignettes, explaining why we failed to replicate Study 1 and 2’s results.

**General Discussion**

**Summary of Results**

Across three studies, we provide some evidence that women labeled as emotional during a disagreement will be perceived as making less legitimate arguments than when men are called emotional, or when an emotion evaluation is absent from the disagreement. In two of three studies, we found evidence that the perceptions of argument validity were not driven by simply being labeled as emotional. Rather, gender coupled with the label led to delegitimization. We did not find support for this pattern in our third study, which used video instead of written stimuli, suggesting that delegitimization may occur when contextual gaps are present, facilitating the use of stereotypes to inform perception.

Mediation analyses suggested that our findings in the first two studies occurred because observers believed the emotional evaluation when it was directed toward women but did not believe it when directed toward men. Specifically, when both women and men were called emotional in identical circumstances, women characters were perceived as more emotional than the men characters. Further, observers perceived the arguments of the labeler calling a woman emotional to be more legitimate than the arguments of either the labeler calling a man emotional or the labeler who did not use an emotion evaluation in disagreements. This finding is consistent with research that has found that gender-emotion stereotypes drive perceptions of women and men’s emotions, regardless of the actual presence of emotion (Robinson et al., 1998), and that men’s displays of emotions, even when they are the same as women’s, are viewed as more appropriate and acceptable (e.g., Warner & Shields, 2007).

Again, however, we failed to replicate this mediation in Study 3—target emotionality did not mediate the relation between condition and target legitimacy. Such a failure limits the strength of what we can conclude from the first two studies. In short, it seems that sometimes an emotional label can result in delegitimization for women in particular, but not at all times. In comparing the samples, stimuli, and limitations of our collection of studies in the next section, we identify potential explanations for Study 3’s unique results as well as boundary conditions influencing the delegitimization process.

**Limitations**

Importantly, our deductions are limited by our small student-based samples and school-based, unvalidated scenarios. Our first two studies were underpowered, despite detecting significant differences by condition. We call for future research to replicate our first two studies with a predetermined, bigger sample size to clarify the presence and size of any observed delegitimization effect, particularly using written vignettes. We suggest testing the boundary conditions of the delegitimization process only after investigating and replicating the foundation set by the current studies. As such, we believe these studies still have value in starting a conversation and providing initial data on the effects of emotional labels, despite the power limitations.

Next, all three studies were conducted with undergraduate students who likely have had little full-time work experience, so we designed our vignettes and measures to reflect school
circumstances that were more familiar to them (e.g., disagreeing over a club budget instead of a client’s budget, or receiving a bad grade instead of receiving a poor job performance review). Our results can only speak to how the emotional label functions in school contexts, rather than in a traditional workplace. Moreover, idiosyncrasies of our various vignettes and the specific contextual factors embedded in them (e.g., the number of days until the project is due) may have caused our pattern of results, as our videos were not validated. However, we believe that they still hold value as an initial investigation of how labels can be employed to invalidate within school circumstances. For example, when other power dynamics (such as those present in a supervisor-subordinate or elected official-constituent relationship) are at play, judgments may differ. Observers may be predisposed to believe those with power, reifying the legitimacy of their higher status.

In addition, across our vignettes, we only used one kind of conflict—that in which two parties have a disagreement about expectations. There are many other kinds of conflict occurring in different contexts—for example, equal conversation partners could be negotiating toward a common goal, or a boss and their subordinate could be discussing the pros and cons of different suggested action plans, and still disagree. Indeed, the fact that our results from all three studies revealed that labelers’ argument were seen as most legitimate when an emotional woman was seen as least legitimate could reflect that participants may have used a zero-sum approach to judging legitimacy in our vignettes. While our scenarios were designed to be ambiguous regarding who is wrong and right, participants may have interpreted them differently, with a clear “winner” and a clear “loser.” Specifically, they may have expected that only one person can be right, and thus the other must be wrong, which could have affected whose argument was seen as legitimate, instead of a perspective in which both arguments had some legitimacy. Such an approach would explain why the labeler was seen as more legitimate when they delegitimized a target through a label, but not when no label was present in Studies 1 and 2. Given these specific circumstances depicted in our vignettes, we do not suggest that the effects of an emotional label in all contexts, every time, would result in the same pattern of results. Future research should investigate the effects of emotional labels in more traditional workplace scenarios with conversation partners who have varying levels of power in varying kinds of conflict as the opening examples in this paper demonstrate.

Our scenarios also omitted additional factors that might interact with situational ambiguity to influence emotionality and legitimacy judgments. For example, when constituents read about an argument between politicians, they normally have more contextual information about the politicians’ track records and may have pre-formed judgements of their competency. Given prior work that demonstrates how men are perceived to have more emotional competency, any indicators of job incompetency might exacerbate judgements based on labels, constituting a double-hit to overall competency and legitimacy. Indeed, prior theorizing argues that cues to competency might further affect perceptions of women leaders’ legitimacy (Vial et al., 2016). We predict that, if we had incorporated incompetency into the vignettes, the differences observed in our pattern of results may have simply been amplified when incorporating other information about job competency. In addition, we did not measure perceptions of other factors that may influence perceived emotionality. For example, an inability to control one’s emotions, at least anecdotally, seems to be an important facet of the stereotype about women’s emotionality and irrationality. We did not measure this perceived self-control, and therefore cannot speak to its role in informing judgements of perceived argument legitimacy. In our set of studies, we aimed to understand how, all other factors assumed equal between two conversation partners, an emotional label could hurt perceived legitimacy alone. We suggest future research incorporate such variables to fully uncover the content and effects of gender stereotypes of emotionality.

Finally, while our third study did not replicate the results of Studies 1 and 2, it did provide further insight into factors that might affect perceptions of argument legitimacy. Delivering the manipulation through a video vignette mimics what it would be like to directly witness the disagreement unfolding. Knowing exactly what was said, how it was said, and the body language that accompanies it may help combat the ambiguity that delegitimization could rely on. As we did not intend to control for and test this kind of difference, however, we cannot speak with confidence to its role and can only hypothesize it as a potential explanation for the inconsistencies in our data. However, given the known role of ambiguity in facilitating discrimination (Barrantes & Eaton, 2018), we believe it a viable avenue for future research and an important factor to consider when attempting to understand the process of delegitimization.

**Future Directions**

In line with our limitations, we believe a pressing area for future research is investigating the role of written versus face-to-face information in making judgments about others’ legitimacy. In our own studies, we observed that our first two experiments, which delivered information in a way that approximated written communication, revealed the predicted pattern of results, whereas our experiment that delivered information in a more “face-to-face” manner did not. It is beyond the scope of this paper to directly compare the two, but we strongly encourage future work to test how the source of information affects the extent to which delegitimization occurs.

Next, while our studies used single third-party observers (namely, the participant) to assess perceptions of two individuals’ legitimacy, a facet of delegitimization involves the
implicit consensus achieved by the group of people assessing the validity of claims. A subsequent study could examine the potentially larger impact that having a third party consisting of several members would have on judgments of legitimacy, especially when no one challenges the evaluation that a woman is being emotional. Prior work has found that when an invalidation is not explicitly or implicitly challenged, the appearance of consensus makes the invalidation seem like objective social reality (Johnson et al., 2006). To this end, future research should examine reactions to third party members who question emotional labels. Indeed, prior research on allyship has found that confrontation of an offender is perceived as more valid when coming from an ally than from a target of bias (Drury & Kaiser, 2014). Investigating the role of group (in)validation and confrontation could, in this way, provide potential solutions and interventions to combatting the delegitimization of women through the emotional label.

Finally, future research should examine how intersecting stereotypes and identities affect the consequences of being labeled as emotional. Although previous evidence suggests that women, overall, are believed to be more emotional than men in mainstream U.S. contexts, (Durik et al., 2006; Shields, 2002), the degree to which this is so, however, may depend on the race of the target and other social identities that women may hold—meaning, it depends upon a woman’s intersectional positioning. According to intersectionality theory, social categories are mutually constitutive (Crenshaw, 1991; Zinn & Dill, 1996). For example, Black women are perceived to be more angry than White women (Durik et al., 2006; Harris-Perry, 2013; Landrine, 1985), whereas White women are perceived as simply emotional (Landrine, 1985). In addition, U.S. stereotypes of East Asian women are that they are reserved and subservient (Hess et al., 2002) and display less intense facial expressions than White women (Matsumoto & Ekman, 1989), suggesting they may not be perceived as being emotional enough.

Not only would future research on this topic combat the intersectional invisibility of these individuals (in which their experiences are not adequately represented at best and misrepresented at worst; Purdie-Vaughns & Eibach, 2008), but it would also clarify the function of stereotypes in delegitimizing women’s arguments. Previous research has found that White participants rate White women as the most prototypical members of their gender group (Coles & Pasek, 2020; Schneider, 2005). With our predominately White sample, when studying a general category like gender we cannot be confident that emotion stereotypes about “women” in general will be an accurate reflection of all women. Our studies are merely a first step in understanding how emotional labels affect legitimacy and are limited in their consideration of intersectionality. Future research should employ an intersectional framework to further examine how emotional labels affect all women, not just the prototypical (White) woman we have studied.

Practice Implications

We believe our findings hold important implications for not only everyday, casual conversations, but also how we navigate national conversations about women more broadly. First, we hope that understanding how an emotional label can delegitimize a woman’s argument can encourage people to be aware of their own subconscious usage of such labels during disagreements, and inspire them to avoid using the label at all. Second, discussion of this phenomenon could be incorporated into workplace training on how gender influences everyday interactions at work. In particular, trainings could address how it is inappropriate and delegitimizing to call a woman emotional in discussions with coworkers at the water cooler or over email, highlighting how calling a woman emotional is one example of the minor instances of bias that accumulate over time to create significant inequities (Chafetz & Valian, 1999). Trainings can help participants identify and change their unconscious behavior to counteract this accumulation of bias.

However, as discussed at the opening of the paper, people also use the label intentionally to delegitimize. Therapists and counselors are likely already aware of abusers’ attempts to delegitimize their victims’ claims by calling them emotional, but through police training, police officers can become aware of this strategy to help inform how they approach calls regarding domestic violence (Sweet, 2019). Testing and increasing awareness of the label’s effects in the justice system and mental health system could potentially protect women better than an unchallenged system. In addition, politicians who hope to delegitimize their women opponents should be challenged when using an emotional label, and news outlets reporting on women in power should review their publications to ensure they are not using biased language. In having these conversations and calling attention to the implications of an emotional label for women, we may be able to start a conversation about the stigma implied in the label, and help women so that the legitimacy of their arguments is assessed based on soundness, rather than stereotypes.

Conclusion

In conclusion, the present research reveals how gender interacts with emotional labels to affect the perceived legitimacy of conversational arguments. Our first two studies found that calling a woman versus a man emotional leads to the lower perceived legitimacy of her arguments. This effect is due to perceptions of women’s higher emotionality in comparison to men, creating a contrast effect in which the less legitimate a labeled woman target’s argument is, the more legitimate the labeler’s argument is. Our third study suggests a potential limitation to this effect—namely, that emotional labels may be most insidious when observed through written means or under ambiguous conditions. While our studies cannot account for all contexts that impact the delegitimization process, we
believe that our studies have important implications for whose arguments we believe are valid, who we put in charge of our clubs and class projects, and more broadly, who we put in charge of our country due to biased stereotypes about the emotionality of men and women.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Florence L. Geis Memorial Dissertation Award, APA Division 35, The Psychology of Women.

ORCID iD
Teresa J. Frasca https://orcid.org/0000-0003-0798-7883

Supplemental Material
Supplement material for this article is available online.

Notes
1. We do not speak to how stereotypes of emotionality may be gendered in cultures in the Global East, nor the effects of an emotional label in these contexts. Research is mixed, with some studies finding no differences in norms of emotional expression and suppression based on culture, and others finding greater pressure against emotional expression—particularly of negative emotions—in East Asian cultures for example (see Han et al., 2022 for a brief review). Our work pulls on and speaks only to the function and effect of Western stereotypes of emotionality (and specifically, those present in the United States). If norms of emotional suppression are even stronger in East Asian cultures, then being labeled as emotional during a disagreement could exacerbate judgements of one’s illegitimacy. As such, future work should investigate if similar patterns of delegitimization happen in other cultures and how the specific context of that culture influences or reveals details about the process of delegitimization.

2. We used G*Power to run a sensitivity analysis, recommended for potentially low-power studies that have already been conducted (Perugini et al., 2018). A 2 × 2 × 2 between-subjects ANOVA with 84 participants across eight groups would be sensitive to effects of η² = 0.43 with 80% power (alpha = .05).

3. We used G*Power to run a sensitivity analysis, recommended for potentially low-power studies that have already been conducted (Perugini et al., 2018). A 2 × 2 × 2 between-subjects ANOVA with 124 participants across eight groups would be sensitive to effects of η² = 0.35 with 80% power (alpha = .05).

4. This change to the control condition (where the labeler does not get to respond to the target’s last statement, versus in Studies 1 and 2 the labeler in the control condition says they disagree or asks the target to see it from their point of view) was made because we were concerned that the labeler’s final statement might influence participants’ interpretation of the emotional label. In Study 3, we wanted to see the effects of a label left without a response or further contextual information from the labeler.

5. To test Hypothesis 5, we conducted 2 (Target Gender: Man or Woman) X 2 (Labeler’s Evaluation: Emotional or Control) between-subjects ANOVAs. For the majority of our outcomes, we did not find support for H5. Contrary to our predictions, target women were perceived to be significantly more likely to be leaders (M = 6.19, SD = 0.14) than target men (M = 5.49, SD = 0.15) regardless of whether or not they were labeled as emotional, F(1, 214) 11.62, p = .001, η² = .05, with no interaction and no main effect of label, all ps > .22. Target women were also assigned significantly higher grades on average (M = 92.56, SD = 0.80) than target men (M = 90.03, SD = 0.86) regardless of the presence of a label, F(1, 214) 4.612, p = .03, η² = 0.02, with no interaction and no main effect of label, all ps > .19.

There was no main effect of target gender, no main effect of label, and no interaction between the two on the extent to which participants would blame the target if their group were to receive a bad grade, all ps > .28. In addition, there were no significant differences in the extent to which participants would be willing to spend extra time for the target, all ps > .19, willing to volunteer to help the target, all ps > .40, nor the extent to which they predicted that the target would be undermined in future group interactions, all ps > .20.

However, we did find support for our hypothesis regarding participants’ perceptions that their ideas would be dismissed in future group interactions. There was no main effect of gender and no interaction, all ps > .46, but there was a main effect of label, such that participants thought their ideas would be more likely to be dismissed and disrespected in future group projects with emotional targets (M = 3.45, SD = 0.13) than with unlabelled targets (M = 3.07, SD = 0.13), F(1, 214) 4.079, p = .05, η² = .02. In addition, a Chi-square test revealed a main effect of gender when examining whether emotional women were likely to be blamed for the group project receiving a bad grade, such that participants were more likely to blame target women than target men for receiving a bad grade, χ²(1) 5.347, p = .02, regardless of the presence of a label, p = .54. It appears that being labeled as emotional does carry some negative implications for targets, but not to the extent nor in as many ways as expected.

References
Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik, & H. Kellerman (Eds.), Theories of emotion (pp. 305–339). Academic Press. https://doi.org/10.1016/b978-0-12-558701-3.50018-1
Barrantes, R., & Eaton, A. A. (2018). Sexual orientation and leadership suitability: How being a gay man affects perceptions of fit in gender-stereotyped positions. Sex Roles, 79(9-10), 549–564. https://doi.org/10.1007/s11199-018-0894-8
Barrett, L. F., & Bliss-Moreau, E. (2009). She’s emotional. He’s having a bad day: Attributional explanations for emotion stereotypes. *Emotion, 9*(5), 649–658. https://doi.org/10.1037/a0016821

Basford, T. E., Offermann, L. R., & Behrend, T. S. (2013). Do you see what I see? Perceptions of gendered microaggressions in the workplace. *Psychology of Women Quarterly, 38*(3), 340–349. https://doi.org/10.1177/0361684313511420

BBC News (2011). David Cameron criticised for “calm down dear” jibe. https://www.bbc.com/news/uk-politics-13211577

Ben-Zee’v, A. (2000). The subtlety of emotions (Reprint ed.). Bradford Books.

Berger, J., Ridgeway, C. L., Fisek, M. H., & Norman, R. Z. (1998). *Gender as a mark of class: The sociology of gender, race, and ethnicity*. Basic Books.

Berger, J., & Uhlmann, E. L. (2008). Can an angry woman get ahead? *Psychological Science, 19*(3), 268–275. https://doi.org/10.1177/0956797607314128

Brescoll, V. L. (2016). Leading with their hearts? How gender stereotypes of emotion lead to biased evaluations of female leaders. *The Leadership Quarterly, 27*(3), 415–428. https://doi.org/10.1016/j.leaqua.2016.02.005

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? *Psychological Science, 19*(3), 268–275. https://doi.org/10.1177/0956797607314128

Coles, S. M., & Pasek, J. (2020). Intersectional invisibility revisited: How group prototypes lead to the erasure and exclusion of Black women. *Translational Issues in Psychological Science, 6*(4), 314–324. https://doi.org/10.1037/tps0000256

Crenshaw, K. (1991). *Mapping the margins: Intersectionality, identity politics, and violence against Women of Color*. Stanford Law Review, 43(6), 1241–1299. https://doi.org/10.2307/1229039

Del Real, J. A. (2016). *Trump, in series of scathing personal attacks, questions Clinton’s mental health*. Washington Post. https://www.washingtonpost.com/news/post-politics/wp/2016/08/06/trump-in-series-of-scathing-personal-attacks-questions-clintons-mental-health/

Drury, B. J., & Kaiser, C. R. (2014). Allies against sexism: The role of men in confronting sexism. *Journal of Social Issues, 70*(4), 637–652. https://doi.org/10.1111/josi.12083

Durik, A. M., Hyde, J. S., Marks, A. C., Roy, A. L., Anaya, D., & Schultz, G. (2006). Ethnicity and gender stereotypes of emotion. *Sex Roles, 54*(7–8), 429–445. https://doi.org/10.1007/s11199-006-9020-4

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111*(1), 3–22. https://doi.org/10.1037/0033-2909.111.1.3

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fischbach, A., Lichtenthaler, P. W., & Horstmann, N. (2015). Leadership and gender stereotyping of emotions. *Journal of Personnel Psychology, 14*(3), 153–162. https://doi.org/10.1027/1866-5888/a000136

Fischer, A. H., & Manstead, A. S. R. (2000). The relation between gender and emotion in different cultures. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 71–94). Cambridge University Press. https://doi.org/10.1017/CBO9780511628191.005

Gross, J. J., John, O. P., & Richards, J. M. (2000). The dissociation of emotion expression from emotion experience: A personality perspective. *Personality and Social Psychology Bulletin, 26*(6), 712–726. https://doi.org/10.1177/0146167200268006

Han, D. E., Park, H. G., An, U. J., Kim, S. E., & Kim, Y. H. (2022). *Emotional restraint is good for men only: The influence of emotional restraint on perceived power and prestige orders*. *American Sociological Review, 63*(3), 379–405. https://doi.org/10.1017/0742265755

Heilman, M. E., & Wallen, A. S. (2010). Wimpy and undeserving of respect: Penalties for men’s gender-inconsistent success. *Journal of Experimental Social Psychology, 46*(4), 664–667. https://doi.org/10.1016/j.jesp.2010.01.008

Hess, U., Beaupre, M. G., & Cheung, N. (2002). Who to whom and why - cultural differences and similarities in the function of smiles. In M. Abel (Ed.), *An empirical reflection on the smile* (pp. 187–216). The Edwin Mellen Press.

Hess, U., David, S., & Hareli, S. (2016). Emotional restraint is good for men only: The influence of emotional restraint on perceptions of competence. *Emotion, 16*(2), 208–213. https://doi.org/10.1037/emo0000125

Johnson, C., Dowd, T. J., & Ridgeway, C. L. (2006). Legitimacy as a social process. *Annual Review of Sociology, 32*(1), 53–78. https://doi.org/10.1146/annurev.soc.32.061604.123101

Landrine, H. (1985). Race x class stereotypes of women. *Sex Roles, 13*(1–2), 65–75. https://doi.org/10.1007/bf00287461

Levi, M., Sacks, A., & Tyler, T. (2009). Conceptualizing legitimacy, measuring legitimating beliefs. *American Behavioral Scientist, 53*(3), 354–375. https://doi.org/10.1177/0002764209337977

MacArthur, H. J. (2019). Beliefs about emotion are tied to beliefs about gender: The case of men’s crying in competitive sports. *Frontiers in Psychology, 10*(2765). https://doi.org/10.3389/fpsyg.2019.02765

MacArthur, H. J., & Shields, S. A. (2015). There’s no crying in baseball, or is there? Male athletes, tears, and masculinity in North America. *Emotion Review, 7*(1), 39–46. https://doi.org/10.1177/1754073914544476

MacArthur, H. J., & Shields, S. A. (2019). How you cry, when you cry, why you cry, and who you are: Responses to adult crying in social contexts. In U. Hess, & S. Hareli (Eds.), *The social nature of emotion expression* (pp. 209–225). Springer. https://doi.org/10.1007/978-3-030-32968-6_1
