# Preregistration: Bayesian Analysis for the Gender, Status, and Emotions Project Study 2

Suzanne Hoogeveen & Julia Haaf

04/08/2021

## Overview

We outline a Bayesian analysis for Study 2 in the Gender, Status, and Emotion project: We will construct various hierarchical models that reflect the predictions from the proposed theories. The evidence for each of these different theories will be quantified by Bayes factor model comparison, following the approach used by Haaf & Rouder (2017) and Rouder, Haaf, Davis-Stober, & Hilgard (2019). In addition, we will adopt a multiverse approach in which we assess evidence for various a priori specified alternative analysis paths related to exclusion criteria (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016).

## The effect

This work involves a large-scale attempted replication of the gender-and-anger backlash effect (Brescoll & Uhlmann, 2008): whereas angry men in a professional context are accorded higher status, angry women are accorded lower status, compared to sad men and sad women, respectively. In addition to attempting to replicate this key effect, various theory-based alternative perspectives are compared. To this end, several moderator variables are added to the design. Furthermore, for Study 2, the materials and study design were crowdsourced across different labs, resulting in 27 different designs (i.e., the original design from Study 1 + 26 conceptual replications). In addition to testing for overall differences between designs, we will assess the influence of several design-related moderators.

### Primary Theoretical Tests

For **status conferral** and **competence**[1] as outcome variables:

1. The *Gender Stereotyping* perspective predicts female managers suffer backlash and decrements in status for expressing anger because they have violated prescriptive norms (Brescoll & Uhlmann, 2008): Anger increases status conferral/competence to male targets, and decreases status conferral/competence to female targets.
2. The *Status Signaling* perspective predicts that anger projects dominance and status for both men and women: Anger increases status conferral/competence to all targets.
3. The *Cultural Change* perspective predicts that anger expressions lead to favorable evaluations of women relative to men, due to exposure to feminist messages: Anger increases status conferral/competence to female targets only, and female targets are generally accorded more status/competence. (Effects are driven by internal motivations and beliefs related to gender inequality, and female evaluators should be more likely to exhibit reverse gender biases, since women are more likely to support #MeToo and feminist beliefs.)
4. The *Study Savviness* perspective similarly predicts that angry women are favorably evaluated, but due to awareness of the study topic, external motivation not to appear prejudiced, and previous research

---

[1]In the remainder of the preregistration document, we will refer to "status conferral" as the dependent variable, though the described predictions and models also apply to "competence".

experience: Anger increases status conferral/competence to female targets only, and female targets are generally accorded more status/competence. (Effects are driven by external motivations related to self-presentation goals to avoid appearing sexist, and should apply to male participants more so than female participants, since the former are more concerned about appearing sexist.)

For **warmth** as outcome variable:

5. The *Anger suppresses warmth* perspective predicts that anger expressions lead to less attributed warmth for both male and female targets: Anger decreases warmth to all targets.

For **out of control** as outcome variable:

1. The *Gender Stereotyping* perspective predicts female managers will be seen as more out of control when expressing anger: Anger increases out-of-control ratings for female targets, compared to non-angry targets and angry male targets.

## Secondary Theoretical Tests

### Cultural Differences

In addition to the main effects of and interaction between target gender and target emotion, we will again assess the influence of culture on these effects. However, given that (1) we don't have the cultural harmony scores for all sites and the (2) we found strong evidence against a cultural differences perspective in Study 1, the cultural dimension will be included in a secondary analysis.

6. The *Cultural Differences* perspective predicts that anger may have positive effects in confrontation (disagreement/assertiveness) oriented cultures but negative effects in more harmony-oriented cultures. Either: interaction such that with increased cultural harmony-score, the positive effect of anger decreases or becomes negative? Or: in low harmony cultures (based on median split or specific value), we expect that anger increases status conferral and in high harmony cultures, anger decreases status conferral (cf. east-west split).

*(Not sure how to translate this into testable predictions; do we simply include a directed interaction between* Suus *the continuous culture thing and emotion condition, or make groups (like in Study 1)?)*

### Main Effects

As in Study 1, we will assess the evidence on the 4 DVs (status, competence, warmth, out of control) for:

1. main effect of target gender
2. main effect of target emotion

### Gender of the Rater

Again, as in Study 1, we will assess the main effect of rater gender on the 4 DVs, as well as the moderating effect of rater gender on the effects of target gender, target emotion, and the interaction.

1. main effect of rater gender
2. interaction between rater gender and target gender
3. interaction between rater gender and target emotion
4. three-way interaction between rater gender, target gender, and target emotion

### Individual Difference Moderators

As in Study 1, we will again assess the evidence for the individual difference measures relevant to distinguishing between perspectives 3 and 4. Following the approach in Study 1, we will compare the evidence in the data for a null-model, an individual differences-only model, the experimental effects model (i.e., primary theoretical tests), a moderation model (i.e., experimental effects + interactions between moderator and target gender, as well as between moderator, target gender, and target emotion), and an unconstrained model (see Study 1 for details).
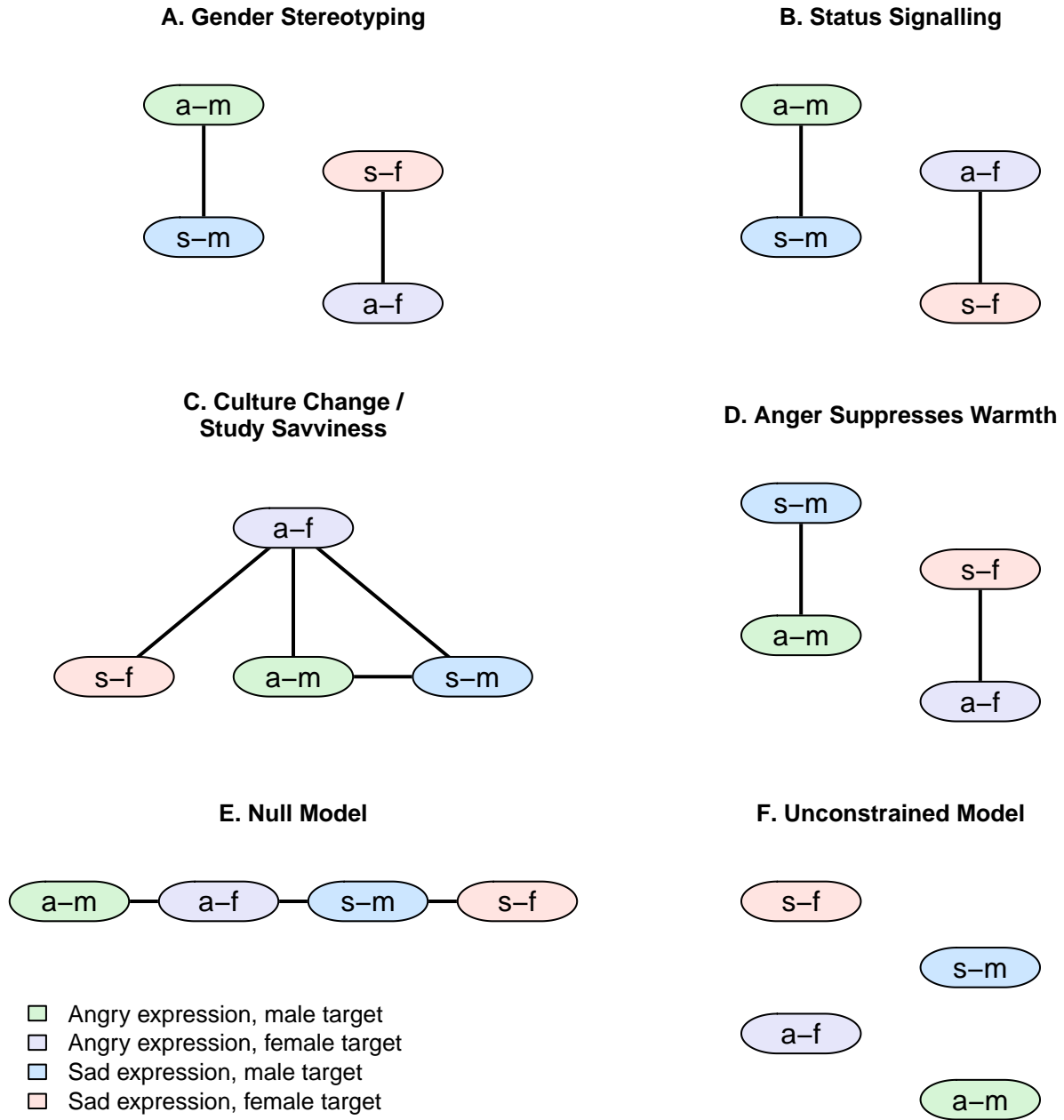
Figure 1: Theoretical positions are captured with order-constraints on cell means. Cells labeled 'a-m' correspond to the condition with an anger expression and a male target, cells labeled 's-m' correspond to the condition with a sadness expression and a male target, cells labeled 'a-f' correspond to the condition with an anger expression and a female target, cells labeled 's-f' correspond to the condition with a sadness expression and a female target. **A.** Gender Stereotyping model. **B.** Status Signalling model. **C.** Cultural Change / Study Savviness model. **D.** Anger Suppresses Warmth model. **E.** Null model. **F.** None-of-the-above alternative model where all orderings of cell means are possible. Note that models **A**-**C** apply to 'status conferral' and 'competence' as outcomes variables, whereas model **D** only applies to 'warmth' as the outcome variable

## Primary Methodological Tests

In addition the testing the theoretical perspectives on gender and emotion expressions, we will also assess to what extent design-related moderators might explain differences between designs.

First, we will investigate the variability in the 2 main effects (target gender and target emotion) + the interaction across study designs and test if we indeed find evidence for substantial variation (i.e., most evidence in favor of a varying effects or unconstrained model). Then, we will identify which moderators affect each of these 3 effects.

Note that we will not conduct any statistical tests for these moderators, but show the patterns by visualizing the effect sizes per category of the moderators.

1. Comparison condition: sadness (designs 1,7,8,10,17,18,23,24,27) or neutral emotion (designs 2-6,9,11-16,19-22,25,26)
2. Format: video (2,15,16,19-26), audio recording (11-14,17,18), comic book (1), or scenario (3-10,27)
3. Manipulation of gender: visual (actors/characters) (1,2,10-26), written (names/pronouns/prefixes/gender) (3-9,27)
4. Manipulation of anger: content of dialogue (1,4-6,9,10,13,14,17,18,21-24,27), tone of delivery (2,3,11,12,15,16,19,20,25,26), or rating on scale (7,8)
5. Type of design: original (27), adapted from existing design (1-8), or novel design developped for this project (9-26)

Second, we will assess the *anger extremity hypothesis a*: moderate levels of anger will lead to more status conferral than extreme anger for both targets. Note: in this analysis, only the designs in which emotion extremity is manipulated will be included (i.e., 7,8,11-26).

## Secondary Methodological Tests

Third, continuous participant ratings of anger extremity, appropriateness, dominance and warmth are investigated as moderators. Specifically:

1. *Anger extremity hypothesis b* expects that more extreme anger, as rated by participants, should be associated with reduced status conferral from anger expression for all targets.

2. *Appropriateness hypothesis* expects that less appropriate anger expressions, as rated by participants, should be associated with reduced status conferral from anger expression for all targets.

3. *Target dominance hypothesis* expects more backlash against angry women in status conferral (i.e., a stronger interaction between target gender and emotion expression) to the extent the anger expression projects dominance.

4. *Target warmth hypothesis* expects more backlash against angry women in status conferral (i.e., a stronger interaction between target gender and emotion expression) to the extent the anger expression projects a lack of warmth.

# Multiverse Analysis

As in Study 1, for the main analyses we use the intent-to-treat approach, in which few to no observations or participants are excluded (Gupta, 2011; McCoy, 2017). In the multiverse, we will further include the following paths:

- Language fluency.

1. Include everyone regardless of years of English experience (or other language in which the survey is administered)

2. Exclude participants with less than 5 years of English experience (or other language in which the survey is administered)

- Manipulation check.

1. Exclude no one based on the manipulation check
2. Exclude participants who did not correctly recall the target gender

- Straightlining.

1. Exclude no one based on pattern of responding
2. Exclude participants who always selected the same option on all items within each of the 5 scales

- Material quality.

1. Exclude no one based on reported quality of materials
2. Exclude participants from video/audio designs who indicated poor video/audio quality

*(How is this measured and what criterion do we use? I assume just a check question "could you hear/see the    Suus recording/video properly?")*

- Mode of administration.

1. Exclude no one based on method of taking the survey
2. Exclude partipants who did not complete the study on a laptop or desktop

- Manipulation failure designs.

1. Include all designs
2. Exclude designs for which there is no compelling evidence (BF=10) for either of or both of the manipulation checks (gender and emotion).

In total, there will be $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$ conditions for the multiverse analysis.

To assess whether we want to do an exclusion or no exclusion at all we suggest to only analyze the level if there is less than 95% overlap with other levels. We assume that otherwise the exclusions will not have a meaningful effect on the results.

For the scales: items will be dropped when alpha < .40 until alpha > .40.

# Model Specification and Predictions

First, we will conduct a manipulation check to validate that manipulation of the targets' emotion expressions. Specifically, we will test whether targets with anger expressions are indeed perceived are more angry than targets with sadness expression and vice versa for sadness. ## General Model Structure

For the main analyses, we will use Bayesian hierarchical modeling with participants nested in countries/labs and in designs. As the random-effects model in Study 1 indicated evidence against variability between countries/labs, and since we're interested in between-desing variability in Study 2, we will use varying intercepts across sites and varying intercepts + varying slopes across designs. We will first construct an unconstrained model that includes all main parameters from the separate theories, which are free to vary in size and direction. In the primary analysis, we will used different ordinal constraints to capture the different theoretical predictions (see below). Bayes factor model comparison will be used to compare the models and determine what theory best predicts the empirical data. This method is based on the work by Haaf, Klaassen, & Rouder (2018).

Let $Y_{ijklm}$ be the status rating of the $i$th lab, the $j$th design, the $k$th participant in the $l$th target gender condition ($l = 1, 2$, for female and male targets, respectively) and the $m$th target emotion condition ($m = 1, 2$, for sadness and anger expressions, respectively). Then

$$Y_{ijklm} \sim N(\alpha_i + \eta_j + x_{1jl}\beta_j + x_{2jm}\gamma_j + x_{3jlm}\theta_j, \sigma^2),$$

where,

- $\alpha_i$ is the baseline rating for the $i$th lab
- $\eta_j$ is the baseline rating for the $j$th design
- $\beta_j$ is the target gender effect for the $j$th design
- $\gamma_j$ is the target emotion effect for the $j$th design
- $\theta_j$ is the gender-by-emotion interaction effect for the $j$th design

and

- $x_{1jl}$ is the indicator for target gender ($l = 1, 2$, for female and male targets, respectively)
- $x_{2jm}$ is the indicator for target emotion ($m = 1, 2$, for sadness/neutral and anger expressions, respectively)
- $x_{3jlm}$ is the indicator for the gender-by-emotion interaction.

## Prior settings

We think small effects in the predicted direction may still be meaningful, especially with regard to gender bias where small biases can accumulate in terms of their consequences over time. We therefore propose to use a scale of 0.25 for the effect of interest. A scale of 0.25 assumes an size effect that is 25% of the sampling noise (standard deviation), which is generally considered a small effect. For the variation between labs in the intercepts, we use a scale of 1.

## Theoretical Predictions

1. Null Model: angry men, sad men, angry women and sad women are all accorded equal status, and there are no differences in status conferral between designs.

From the general model, the null-model is adjusted as follows:

$$Y_{ijklm} \sim N(\alpha_i, \sigma^2).$$

where $\alpha_i$ is the baseline status conferral rating for $i$th lab.

2. Designs Baseline Model: angry men, sad men, angry women and sad women are all accorded equal status, but status conferral differs between designs.

From the general model, the baseline-model is adjusted as follows:

$$Y_{ijklm} \sim N(\alpha_i + \eta_j, \sigma^2).$$

3. Gender Stereotyping: while angry men are accorded higher status than sad men, angry women are accorded lower status than sad women.

$$Y_{ijklm} \sim N(\alpha_i + \eta_j + x_{1jl}\beta_j + x_{2jm}\gamma_j + x_{3jlm}\theta_j, \sigma^2).$$

Effect coding is used to quantify the different conditions. The ordinal constraints based on the theoretical perspective are put on the cell means, rather than on the parameters. Specifically, the cell mean of the angry men condition needs to be higher than the sad men condition and the cell mean of the sad women condition needs to be higher than the cell mean of the angry women condition. Cell means are calculated based using the estimated parameters; for the sad women condition, for instance, this results in the following: $Y_{\cdot\cdot11} = -1/2\beta - 1/2\gamma + 1/2\theta$.

To satisfy the theoretical predictions, the following inequality constraints have to hold:

$$Y_{\cdot\cdot11} > Y_{\cdot\cdot12}$$
$$Y_{\cdot\cdot22} > Y_{\cdot\cdot21}$$

4. Status Signaling: angry men are accorded higher status than sad men and angry women are accorded higher status than sad women.

The status signaling perspective applies the same model and parameters as the gender stereotyping model. Here, the cell mean of the angry men condition needs to be higher than the sad men condition and the cell mean of the angry women condition needs to be higher than the cell mean of the sad women condition. To meet this condition, the following inequality constraints have to hold:

$$Y_{..12} > Y_{..11}$$
$$Y_{..22} > Y_{..21}$$

5. Culture Change/ Study Savviness: angry women are accorded more status than sad women, angry men and sad men. There is no effect of emotion for men.

$$Y_{ijkl} \sim N(\alpha_i + \eta_j + x_{4kl}\delta + x_{5kl}\eta, \sigma^2),$$

where parameter $\delta$ is the effect of emotion for female targets, parameter $\eta$ is the effect of angry women versus men (both sad and angry). The indicator variables are $x_{4lm}$ (-1/2 for sad female targets and 1/2 for angry female targets, and 0 for men) and $x_{5lm}$ (-1/3 for men, 2/3 for angry women and 0 for sad women).

The theoretical perspective entails that the cell mean of the angry women condition needs to be higher than the sad women condition and higher than the cell means of the angry men and the sad men conditions. To meet this condition, we will put inequality constraints on the parameters:

$$\delta > 0$$
$$\eta > 0$$

6. Anger suppresses warmth: sad men are considered warmer than angry men and sad women considered warmer than angry women.

The anger suppresses warmth model applies the same model and parameters as the status signalling model, but constrains the emotion parameters in exactly the opposite direction:

$$Y_{..12} < Y_{..11}$$
$$Y_{..22} < Y_{..21}$$

7. Cultural Differences: in confrontation-oriented cultures, angry men and women are accorded more status than sad men and women, while in harmony-oriented cultures, sad men and women are accorded more status than angry men and women.
   This model builds on the model for perspectives 2 and 3 and is extended by parameters for the main effect and interactions of culture (with $m = 1, 2$, for Western and Eastern cultures, respectively):

$$Y_{ijklm} \sim N(\alpha_i + x_{1k}\beta + x_{2l}\gamma + x_{3kl}\theta + x_{6m}\zeta + x_{7km}\xi + x_{8lm}\upsilon + x_{9klm}\omega, \sigma^2),$$

*(adjust this model!)*                                                                                      *Suus*

where the additional parameter $\zeta$ is the effect of culture, parameter $\xi$ is the culture-by-gender interaction effect, parameter $\zeta$ is the culture-by-emotion interaction effect, and parameter $\omega$ is the culture-by-gender-by-emotion three-way interaction. The additional indicator variables are $x_{6m}$ (-1/2 for Western cultures and 1/2 for Eastern cultures), $x_{7km}$ (-1/2 for men in Western cultures and women in Eastern cultures, and 1/2 for women in Western cultures and men in Eastern cultures), $x_{8lm}$ (-1/2 for anger in Western cultures and sadness in Eastern cultures, and 1/2 for sadness in Western cultures and anger in Eastern cultures), and $x_{9klm}$ (-1/2 for sad women and angry men in Western cultures and angry women and sad men in Eastern cultures, and 1/2 for angry women and sad men in Western cultures and sad women and angry men in Eastern cultures).

The ordinal constraints are again put on the cell means, rather than on the parameters. Specifically, the cell mean of the angry men condition needs to be higher than the sad men condition and the cell mean of the sad women condition needs to be higher than the cell mean of the angry women condition. Cell means are calculated using the estimated parameters; for the sad women in Western cultures condition, for instance, this results in the following: $Y_{..111} = -1/2\beta - 1/2\gamma + 1/2\theta - 1/2\zeta + 1/2\xi + 1/2\upsilon - 1/2\omega$.

To satisfy the theoretical ordinal constraints, the following has to hold:

$$Y_{..121} > Y_{..111}$$
$$Y_{..221} > Y_{..211}$$
$$Y_{..112} > Y_{..122}$$
$$Y_{..212} > Y_{..222}$$

8. Unconstrained model: all effects are included, without any ordinal constraints (this is again the genenal model).

$$Y_{ijklm} \sim N(\alpha_i + \eta_j + x_{1jl}\beta_j + x_{2jm}\gamma_j + x_{3jlm}\theta_j, \sigma^2),$$

Note that in contrast to Study 1, we do not include the culture dimension in the unconstrained model, as culture is not relevant for the primary theoretical tests.

# References

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an Angry Woman Get Ahead?: Status Conferral, Gender, and Expression of Emotion in the Workplace. *Psychological Science*, *19*, 268–275. doi:10.1111/j.1467-9280.2008.02079.x

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*, 109–112. doi:10.4103/2229-3485.83221

Haaf, J. M., Klaassen, F., & Rouder, J. N. (2018). Capturing Ordinal Theoretical Constraint in Psychological Science. *PsyArXiv.* doi:10.31234/osf.io/a4xu9

Haaf, J. M., & Rouder, J. N. (2017). Developing constraint in Bayesian mixed models. *Psychological Methods*, *22*, 779–798. doi:10.31234/osf.io/ktjnq

McCoy, C. E. (2017). Understanding the Intention-to-treat Principle in Randomized Controlled Trials. *Western Journal of Emergency Medicine*, *18*, 1075–1078. doi:10.5811/westjem.2017.8.35985

Rouder, J. N., Haaf, J. M., Davis-Stober, C. P., & Hilgard, J. (2019). Beyond overall effects: A Bayesian approach to finding constraints in meta-analysis. *Psychological Methods*, *24*, 606–621. doi:10.1037/met0000216

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science*, *11*, 702–712. doi:10.1177/1745691616658637