

Supplemental Material

Table of contents

Study 1	2
Study 2	9
Study 3	11
Study 4	13
Additional References.....	21

Study 1

Methods

Participants. Participants were recruited on campuses of two German universities. The final sample consisted of 95 participants (63 women, 32 men; $M_{age} = 23.07$, $SD_{age} = 2.95$, 98.9% German).

Because this was a lengthy and complicated lab study and we tested a novel effect, we aimed to recruit at least 50 participants per cell. Using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), we determined with a sensitivity analysis that, to reach 80% power, a sample of $N = 95$ would require an effect of $g_{Hedges} = 0.58$.

Procedure.

Participants were told that they would be taking part in an investigation of "the influence of the spatial distance between team members on their performance", in a team of three. As part of the cover story, participants learned from the experimenter that their team had randomly been placed into one of three conditions. In their condition, team members were in different rooms of the same building with different experimenters and could communicate with each other through the computer. The other two fictional conditions comprised all members being in the same room and all members working from home. To prevent suspicion, participants were already told at the time of recruitment that they would be either asked to come to the laboratory or complete the study from home. The experimenter rehearsed their role based on a standardized script, which can be obtained in German from the first author.

Participants arrived at the laboratory individually. As part of the cover story, they learned from the experimenter that the other participants were in different rooms and that they could communicate with each other through the computer. In reality, the other team members were fictitious and all of their messages had been programmed in the survey software. Participants were told that they would randomly be assigned the role of either the

manager or one of the two employees. The experimenter explained that all team members would individually work on two rounds of eight problem-solving tasks of high and low difficulty. The tasks were mathematical and verbal problem-solving tasks. Participants had 60 seconds to complete each task. To prevent potential feelings of frustration, the experimenter added that only very few participants had been able to solve the difficult tasks.

The task difficulties would be distributed by the manager. We designed tasks that were either easy to solve or unsolvable within the given time frame. For every correctly solved task, participants would collect tickets for a lottery between team members, where they could win 5€ in addition to the 5€ every member received as compensation for participation. Therefore, the more tasks they solved correctly, the more tickets they received and the higher their chances became to beat the other two team members in the lottery. The lottery was implemented to increase participants' interest in a fair task difficulties distribution. Because tangible resources were at stake, participants were given a reason for potential protesting.

After participants gave their informed consent and completed a demographic questionnaire, the experimenter asked them to follow instructions on the computer screen where they would learn about their assigned role and went into the second room of the laboratory.

All participants were assigned to a low power role as “employee #1”. As an expression of high power, the manager could distribute task difficulties and give performance feedback to the employees (the power induction procedure was inspired by Anderson & Berdahl, 2002). After they learned about their role, participants were presented with the manager’s task difficulty distribution for the first round in which he assigned the tasks unfairly. Out of 12 easy and 12 difficult tasks, the manager assigned eight easy and zero difficult tasks to himself and two easy and six difficult tasks to each employee. Participants

were next directed to the tasks. Upon completion of the last task, the program displayed a table, which informed them how many tasks everyone solved and their current chances to win the lottery. Independent of the participants' performance, the table always displayed that each employee had solved two tasks correctly and that the manager had succeeded in all eight. Next, in disguise of performance feedback, participants received a message from the manager that contained *help*: "Hey, you seem to have trouble with the tasks, they were probably too difficult...Let me help you – I'll give you more easy ones in the next round".

The help offer was followed by the *experimental manipulation*. Participants were randomly assigned to either the experimental condition, in which they had to choose one out of three predetermined expressions of thanks or the control condition where they chose one out of three neutral responses. In consideration of our experimental manipulation, the experimenter told the participants that they could only address the manager through predetermined answers to prevent the employees from influencing his distribution of task difficulties. The neutral responses were "I have received the message", "I have read the message" and "I have received the information". Expressions of thanks were ascending in the strength of expressiveness of thanks: "Thank you", "Thank you very much", "Great, thank you very much". We employed the ascending order as an attempt to minimize reactance. We established in a pretest with a separate sample ($N = 50$) that all three thankful responses were perceived as significantly more expressive of thanks than the three neutral responses (detailed analyses can be obtained from the first author). Within both conditions, the respective three responses were presented side by side in a randomized order.

Before the second round, participants saw a (programmed) message saying that the experiment was aborted due to an internal error and were forwarded to the last page which instructed them to contact the experimenter. The experimenter was prepared to inform them that she had just received a text from the other two experimenters, reporting the same

problem. She explained that the study was in its starting phase, that this error had happened before, and that participants would still keep their lottery tickets. However, because the study needed improvement, participants could fill out an evaluation questionnaire, in which they could state what they would do to change the study. This questionnaire assessed *protest intentions* against the manager, on behalf of the employees. To increase the realism of the scenario, the instruction stated that the proposed actions were based on criticism about managers and employees that previous participants had left in the comments window below.

Upon completion of the “evaluation questionnaire”, participants received a message from “employee #2” who complained that the manager had been unfair. They planned to confront him in their next message and encouraged the participant to do the same. Hereby we aimed to set a social norm, which should facilitate protest against the manager. Participants’ messages to the manager formed the *protest behavior* measure.

Next, participants filled out a paper-pencil questionnaire containing the manipulation and power and intergroup - induction checks. Finally, they were debriefed and reimbursed with 10€ for participation.

Measures.

Protest intentions. Protest intentions on behalf of the employees were measured with nine items ($\alpha = .81$; 7-point scales, 1 = *would definitely not participate in*, 7 = *would definitely participate in*). Actions were:

1. ...demanding that the manager hands over lottery tickets to the employees.
2. ...not taking action against the manager.
3. ...calling out the manager together with the other employee.
4. ...leaving the manager alone.
5. ...telling the manager that his decisions were unfair.

6. ...demanding that the manager hands over lottery tickets to me and the other employee.
7. ...demanding that the manager is punished for his unjust distribution of tasks.
8. ... advocating for all people, who participate in the experiment in the future, to have equal power and influence over decisions.
9. ... demanding that the tasks are randomly assigned by the computer.

One item was excluded because of extreme skewness and kurtosis.

Protest behavior. Two research assistants who were blind to the study hypotheses coded the messages to the manager as “protest”, “no protest” or “ambivalent”. The “ambivalent” category was later on merged with the “no protest” category, because ambivalent messages often started with an expression of discontent, but ended with a justification of the manager's behavior. Our measure of protest can, therefore, be considered conservative. Interrater reliability was $\kappa = .72$ and discrepancies were resolved by the first author.

Manipulation, power and intergroup induction checks. As a manipulation check, we administered the item “I thanked the manager for his feedback” to which participants could respond “yes” or “no”.

To establish that the relative power induction was effective, we asked participants to rate on 7-point scales (from 1 = *none at all* to 7 = *very much*) how much power they had over the manager and how much power the manager had over them.

To establish that the induction of an intergroup context was effective (i.e. the other employee is perceived as a member of the ingroup and the manager as a member of the outgroup), we asked participants to indicate how they experienced the interaction with the other employee by choosing whether they experienced it as an interaction between a) two

members of the same group, b) two members of two different groups, or c) two individual students.

Results and Discussion.

Preliminary analyses.

Expression of thanks. Gabriel's pairwise comparisons test showed no significant mean differences in protest intentions between the expressions of thanks ($ps > .662$). Therefore, we aggregated all expressions of thanks into one condition. However, the means for two neutral expressions significantly differed in protest intentions ($p = .039$). Because these cells could neither be aggregated nor treated as separate conditions (due to small cell sizes, 12 and 17), we decided to exclude participants in the smaller cell ("I have read the message"). We excluded the smaller cell to keep as many participants in the sample as possible and also because the responses in the other two cells were more similar in content, both being about having received the message. The remaining two neutral expressions were collapsed and the sample size was reduced to 95.

Protest behavior. Fifty-eight (59.2%) participants expressed protest against the manager (e.g., "Hello, too bad that your task assignment wasn't very fair") and 25 (30.9%) did not express protest (e.g., "Hello dear manager, I'm going home now. See you tomorrow!"). Responses from three participants could not be categorized and therefore the three participants were excluded. Twelve participants did not write to the manager. Because values were not missing completely at random (Little's MCAR test: $X^2(2) = 8.16, p = .017$), nonresponse was handled with full-information maximum likelihood in *Mplus*.

Manipulation, power and intergroup induction checks. Four participants had missing values. A X^2 test showed a significant association between the condition and the manipulation check variable, confirming the different nature of the two conditions, ($X^2(1) = 66.92, p < .001$).

A paired samples *t*-test showed that the manager was perceived as having significantly more power over the participant ($M = 5.51, SD = 1.75$) than the participant over the manager ($M = 1.62, SD = 1.05$), $t(94) = 16.13, p < .001$).

A significant X^2 test showed that the majority of participants (69.47%) perceived the interaction with the employee as an interaction between a) two members of the same group compared to b) two members of two different groups (9.47%) and c) two individual students (21.05%), $X^2(1) = 14.41, p < .001$ (b and c are collapsed).

A X^2 test approached significance in showing that the majority of participants (60%) perceived the interaction with the manager as an interaction between b) two members of two different groups compared to a) two members of the same group (17.89%) and c) two individual students (22.11%), $X^2(1) = 3.80, p = .051$ (a and c are collapsed).

Expression of thanks and protest intentions. The two experimental conditions were coded as 0 = no thanks ($n = 42$) and 1 = thanks ($n = 53$). Welch's *t*-test showed a significant effect of expression of thanks on protest intentions, $t(92.99) = 2.03, p = .045, 95\% \text{ CI } [0.01, 0.91], g_{Hedges} = .42, 95\% \text{ CI } [0.01, 0.83]$. In line with expectations, willingness to protest was lower for participants who expressed thanks ($M = 3.89, SD = 1.24$) than for participants who did not express thanks ($M = 4.35, SD = 0.97$).

Expression of thanks and protest behavior. We conducted a mediation analysis in *Mplus 7.3* (Muthén & Muthén, 1998-2012, 10,000 bootstrap samples for bias corrected bootstrap confidence intervals) in order to test for an indirect effect of thanking on protest behavior mediated by protest intentions. There was no significant total effect of thanks on protest behavior ($b = -0.77, SE = 0.64, p = .233, 95\% \text{ CI } [-2.05, 0.42]$). However, as expected, expressing thanks negatively predicted protest intentions ($b = -0.46, SE = 0.22, p = .040, 95\% \text{ CI } [-0.90, -0.02]$) which in turn positively predicted protest behavior ($b = 0.83, SE = 0.32, p = .009, 95\% \text{ CI } [0.301, 1.52]$). The analysis was confirmed by a significant indirect

effect ($b = -0.38$, $SE = 0.25$, 95% CI [-1.06, -0.04]). The direct effect of the expression of thanks on protest behavior was non-significant ($b = -0.38$, $SE = 0.60$, 95% CI [-1.55, 0.74]).

The present results provide the first evidence that expressions of thanks to a member of a high power group for help inhibit protest intentions on behalf of their ingroup in members of a low power group. There was no main effect on protest behavior, maybe because some participants already expressed protest by indicating protest intentions. However, expressions of thanks indirectly affected behavior through protest intentions.

Study 2

Study 2a

Method.

Participants. Eight participants who indicated that they were non-students or already possessed a graduate degree were excluded from analyses. We expected that an undergraduate scenario would be less relevant to them and that some of them could be already teaching. We excluded another participant who gave the same response to many consecutive items. The data revealed similar outcomes of analyses when these participants were included.

Measures.

Protest intentions. Scale: “would definitely not participate in ...“ (1) to “would definitely participate in...” (7).

1. ...not taking action against the professor.
2. ...leaving the professor alone.
3. ...calling out the professor together with the presentation group.
4. ...jointly telling the professor that his grading was unfair.
5. ...jointly complaining about the professor to the student council.
6. ...jointly describing the incident to the dean.

7. ...jointly addressing the incident at the faculty board meeting.
8. ...jointly writing an anonymous letter to the head of the department complaining about the professor.

Four items were excluded because they had low communalities and to improve reliability.

Study 2b

Method.

Participants. Thirteen participants who indicated that they were non-students or already possessed a graduate degree were excluded from analyses. We expected that an undergraduate scenario would be less relevant to them and that some of them could be already teaching. We further excluded one participant who had 48% missing values on the main variables and another three who gave the same response to many consecutive items. Three participants who indicated previous participation in a similar study were also excluded. The data revealed similar outcomes of analyses when these participants were included.

Measures.

Forgiveness. Scale: 1 = *not at all* to 7 = *very much*.

1. I forgive the professor.
2. I forgive the professor, no matter whether his behavior was right or wrong.
3. I excuse the professor's behavior.

Results Study 2

Alternative causal direction of the mediation model. In order to test the reverse causal direction of the mediation model with forgiveness as the independent variable, expression of thanks as mediator and protest intentions as the dependent variable, we conducted a mediation analysis in *Mplus 7.3* (Muthén & Muthén, 1998-2012, 10,000 bootstrap samples for bias corrected bootstrap confidence intervals). We collapsed the data from Studies 2a and 2b to have a sufficient sample size for using structural equation

modeling ($N = 389$). Forgiveness positively predicted expressing thanks ($b = 0.40$, $SE = 0.06$, $p < .001$, 95% CI [0.28, 0.52]) but expressing thanks did not predict protest intentions ($b = -0.07$, $SE = 0.07$, $p = .309$, 95% CI [-0.22, 0.06]). The indirect effect was not significant ($b = -0.03$, $SE = 0.03$, $p = .317$, 95% CI [-0.09, 0.02]). The direct effect of the expression of thanks on protest behavior was significant ($b = -0.27$, $SE = 0.08$, $p < .001$, 95% CI [-0.43, -0.14]). In order to also confirm the hypothesized direction of the indirect effect for the aggregated sample (IV: expression of thanks, mediator: forgiveness, DV: protest intentions), using structural equation modeling, we conducted a mediation analysis in *Mplus 7.3* (Muthén & Muthén, 1998-2012, 10,000 bootstrap samples for bias corrected bootstrap confidence intervals). Expressing thanks positively predicted forgiveness ($b = 1.03$, $SE = 0.16$, $p < .001$, 95% CI [0.73, 1.35]), which in turn negatively predicted protest intentions ($b = -0.29$, $SE = 0.11$, $p = .008$, 95% CI [-0.54, -0.11]). The analysis was confirmed by a significant indirect effect ($b = -0.30$, $SE = 0.12$, $p = .013$, 95% CI [-0.61, -0.11]). The direct effect of the expression of thanks on protest behavior was non-significant ($b = -0.11$, $SE = 0.17$, $p = .540$, 95% CI [-.55, 0.15]).

Study 3

Method

Participants. The initial sample consisted of 283 female MTurk workers who reside in the United States. We excluded thirteen participants, who failed to correctly copy the response that formed the experimental manipulation (see measures). Two participants responded carelessly and were also excluded. We embedded an attention test-item in the protest intentions items: “telling my colleague that I am disappointed with his behavior. Please ignore this question and choose number 4 as your response.” Twenty participants who failed the attention check were excluded, reducing the final sample to 248 participants.

Measures.

Forgiveness. Scale: 1 = *not at all* to 7 = *very much*.

1. ...I forgive my colleague.
2. ... I forgive my colleague, no matter whether his behavior was right or wrong.
3. ... I excuse my colleague's behavior.

Gender-work-specific system justification. Scale: 1 = *not at all* to 7 = *very much*.

1. ... The division of labor between men and women in the workplace generally operates as it should.
2. ... Gender roles in the workplace need to be radically restructured.
3. ... Work environments are set up so that men and women usually get what they deserve.
4. ... Sexism at work is getting worse every year.
5. ... In general, work relations between men and women are fair.

The three other items from the *gender-specific system justification* measure from Jost and Kay (2005) could not be rephrased for a work context.

Protest intentions. Scale: "I would definitely not engage in ..." (1) to "I would definitely engage in..." (7).

1. ...not taking action against my colleague.
2. ...leaving my colleague alone.
3. ...calling out my colleague.
4. ... telling my colleague that his behavior was unfair.
5. ... writing an email to my colleague explaining why his behavior was not okay.
6. ... complaining about my colleague to the women's representative.
7. ... describing the incident to the boss.
8. ... reporting the incident to Human Resources.

9. ... writing an anonymous letter to Human Resources complaining about my colleague.
10. ...not making an issue out of the incident.
11. ...criticizing my colleague in the next anonymous performance review.

Results

Serial mediation analysis. We conducted a serial mediation analysis in PROCESS (Hayes, 2013, Model 6, 10000 bootstrap samples for bias corrected bootstrap 95% confidence intervals) to examine the indirect effect through forgiveness followed by GWSJ. Figure 1 displays the regression coefficients. The specific indirect effect through forgiveness and GWSJ reached significance ($b = -.01$, $SE = .01$, 95% CI [-0.05, -0.001]). The specific indirect effect through forgiveness alone also reached significance ($b = -.13$, $SE = .07$, 95% CI [-0.29, -0.02]), while the specific indirect effect through GWSJ did not ($b = .04$, $SE = .05$, 95% CI [-0.06, 0.14]). The contrast of the significant indirect effects (forgiveness minus forgiveness → GWSJ) was significant ($b = -.12$, $SE = .06$, 95% CI [-0.27, -0.02]), with the specific indirect effect through forgiveness being the stronger effect. The direct effect was not significant ($b = 0.07$, $SE = .13$, $p = .615$, 95% CI [-0.19, 0.32]).

Post hoc parallel mediation analysis. We conducted a parallel mediation analysis in PROCESS (Hayes, 2013, Model 4, 10000 bootstrap samples for bias corrected bootstrap 95% confidence intervals) in order to explore whether forgiveness and GWSJ will mediate the effect simultaneously. The indirect effect was not significant ($b = -.11$, $SE = .09$, 95% CI [-0.30, 0.05]). The direct effect was not significant ($b = 0.07$, $SE = .13$, 95% CI [-0.19, 0.32]).

Study 4

Method

Participants. The initial sample consisted of 103 female undergraduate psychology students. We excluded four participants, who had dropped out in the follow-up and four who

indicated on a check-item that they did not take participation in the study seriously or did not answer truthfully. We also excluded one who did not believe that the other two students existed (as can be inferred from their open format answers) and one participant who accidentally participated in the first part of the study twice. This reduced the final sample to 93 participants. Three participants failed the manipulation check and three participants were undergraduate cognitive science students, but because the pattern of results did not change when including those participants, we kept them in the final sample.

Procedure. The study was advertised to undergraduate psychology students as supposedly assessing evaluations of our current application procedure for research assistants at our social psychology department. While participating in the study, students could additionally apply for a research assistant position at our lab that would open up in a year (we chose a distal time point to minimize competition). Students could really apply and we have contacted those students who indicated their interest in the position. As in Study 3, we targeted women as the lower power group because we could recruit female students feasibly without having to reveal that gender is the targeted dimension.

Participants came into the lab and were supposedly interacting with two other students through the computer to ensure anonymity. These two participants were apparently in the adjoining rooms with other experimenters, as indicated by signs on the doors. To increase credibility that the other participants existed, they were considered while scheduling lab appointments with the participants. Additionally, at the start of a session, the experimenter (who was blind to the hypotheses) would pretend to communicate with the other two experimenters to synchronize the start of the tasks. We pre-programmed the other students' responses.

The study also ostensibly tested a new procedure for the application process, where applicants were preselected by an independent evaluation team of two, by means of

individual and group performance and mutual evaluations among the participants.

Supposedly, the participants' test results and statements, and their evaluations of the other two participants were immediately submitted to one member of the evaluation team. To justify the follow-up, participants were told that, in about a week, they will receive a questionnaire by email requesting them to provide further details that will be sent to a second member of the evaluation team for independent pre-selection. We asked participants to take part in the study even if they were not interested in applying for the position because their opinion about the other participants was important for the preselection.

After participants gave their consent to participate and answered demographic questions, they received a female fantasy name as a pseudonym and learned the pseudonyms of the other two students. Fantasy names and gender symbols indicated the participants' gender. Participants saw that one student was a female psychology student called "Jileen" (female, therefore same low power group member) and the other a male cognitive sciences student called "Khyron" (male, therefore high power group member). We chose a cognitive sciences student instead of a psychology student because trial runs showed that it did not seem credible that a male psychology student would say something this sexist (→ transgression), because male students are the gender minority in psychology and because it was likely that participants knew most of them.

First, participants read that one important goal of the study was to improve the selection process for research assistants in social psychology. For this, they read about the previous application procedure and were asked to make suggestions for improvements to the procedure. These would be considered in the ongoing procedure. Participants sent their suggestions and could see what the other students suggested. We implemented time intervals so that participants would get the impression that the others were really writing a text. In his suggestion, the male participant made a sexist remark. He suggested a quota for male

research assistants because psychology was female-dominated and men were beneficial to psychology, given that they won all the Nobel prizes. This sexist remark indicated the high power group members' transgression. He also wrote that he will apply for the position as a research assistant.

This was followed by the job advertisement, which also informed about the kind of tasks they will engage in as research assistants (e.g., translation tasks). Fittingly, next, the group had to jointly translate an item from English to German. The other two participants were apparently randomly assigned to the roles of the translators and the participant had to evaluate their translation. While the other participants worked on the translation, the participant completed another filler task (where they had to come up with a solution to how they, as research assistants, would deal with a situation, in which a professor and a Ph.D. student had given them an urgent task at the same time).

After participants received and rated the other students' translation, they could see the chat between them showing the process of the translation. We instructed participants to read the chat carefully so that it is possible for them to evaluate the participants later. In the chat, the male student wrote that he will be a gentleman and transfer his course credits to the "ladies" because he did not need them and was only participating to apply for the position. He had checked with this experimenter who said that this was possible and shared out his credit points between the female participants. This formed the high power group help. Participants then read in the chat that the female student expressed thanks for the male student's help. This was meant to incorporate naturalistic demands to increase the salience of gratitude norms, which should motivate participants to express thanks.

Then, we experimentally manipulated the opportunity to express thanks: Two-thirds of participants were given a line of communication in response to that chat where they could briefly write to the group if they wished to. The other third could not respond to the chat. We

used a 2:1 ratio because we expected that some participants will not express thanks. Thus, this created another control group condition, where participants did not express thanks although they could.

Next, participants were asked to indicate what they felt when thinking about the other two participants. This measured whether participants felt forgiveness. To measure system justification, participants responded to statements regarding university life. Instructions stated that this information would not be sent to the evaluators, and merely served to capture an opinion. Then we implemented the manipulation check.

Next, participants could state their assessment of the other participants regarding their suitability as research assistants. Participants' statements were supposedly immediately sent to the first member of the independent evaluation team. These statements included protest behavior measures among filler items. Because the male student was interested in applying for the research assistant position, whether participants protested or not mattered in real life and supposedly jeopardized the success of his application. After participants rated the current application procedure, they could directly write to the other two students, which formed another protest behavior measure. Afterward, participants indicated their interest in applying for the position as a research assistant and whether they took participation in the study seriously and answered everything truthfully. Then, their responses were sent to the first independent evaluator. We asked participants not to exchange information about the study with their fellow students to ensure fairness and anonymity of the application process, until a specific date.

Participants responded to the same protest measures online about one week later ($M_{days} = 6.81$, $SD_{days} = 1.45$). They learned, that the male, but not the female participant has applied for the research assistant position so that they only had to evaluate the male participant. Again, participants indicated whether they responded truthfully and their answers

were supposedly sent to the second independent evaluator. After the follow-up, participants were fully debriefed and once again asked to indicate their consent. In addition to the credit points that corresponded with the duration of participation, they were compensated with those credit points, which they had “received” from the male student.

Measures.

Felt forgiveness. Feelings of forgiveness were assessed with three items, e.g., “Forgiveness”, ($\alpha = .81$, 7-point scales, 1 = *not at all* to 7 = *very much*).

Gender-academia-specific system justification (GASJ). We adjusted the *gender-specific system justification* measure from Study 3 to the university context. Four items measured gender-academia-specific system justification, e.g., “In general, gender relations at universities are fair.” ($\alpha = .81$).

Manipulation check. Before assessing protest behaviors the first time, we asked participants whether they yet had a chance to write to the other students to which they could respond “yes” or “no”.

Results

Manipulation check. A X^2 test showed a significant association between the condition and the variable indicating the opportunity to express thanks, confirming the different nature of the two conditions, ($X^2(1) = 80.14, p < .001$).

Expression of thanks vs. no opportunity to express thanks. Welch’s t -test showed a significant effect of expression of thanks on protest behavior for both measurement points, $t_{T1}(44.56) = 2.86, p = .006, 95\% \text{ CI } [0.05, 0.30], g_{HedgesT1} = 0.76, 95\% \text{ CI } [0.27, 1.26], t_{T2}(42.14) = 2.82, p = .007, 95\% \text{ CI } [0.04, 0.26], g_{HedgesT2} = 0.66, 95\% \text{ CI } [0.17, 1.15]$. In line with expectations, participants who expressed thanks showed less protest behavior ($M_{T1} = 0.09, SD_{T1} = 0.16; M_{T2} = 0.10, SD_{T2} = 0.13$) than participants who did not express thanks ($M_{T1} = 0.28, SD_{T1} = 0.32; M_{T2} = 0.26, SD_{T2} = 0.28$).

Expression of thanks vs. no thanks when opportunity was given. An independent samples *t*-test showed a significant effect of expression of thanks on protest behavior for both measurement points, $t_{T1}(59) = 3.68, p = .001, 95\% \text{ CI } [0.09, 0.29], g_{HedgesT1} = 0.96, 95\% \text{ CI } [0.43, 1.50], t_{T2}(59) = 2.11, p = .039, 95\% \text{ CI } [0.005, 0.17], g_{HedgesT2} = 0.55, 95\% \text{ CI } [0.03, 1.07]$. In line with expectations, participants who expressed thanks showed less protest behavior ($M_{T1} = 0.09, SD_{T1} = 0.16; M_{T2} = 0.10, SD_{T2} = 0.13$) than participants who did not express thanks ($M_{T1} = 0.28, SD_{T1} = 0.24; M_{T2} = 0.19, SD_{T2} = 0.20$).

Serial mediation analysis. We conducted a serial mediation analysis in PROCESS (Hayes, 2013, Model 6, 10000 bootstrap samples for bias corrected bootstrap 95% confidence intervals) to examine the indirect effect through forgiveness followed by gender-academia-specific system justification (GASJ). The positive effect of expressing thanks on forgiveness approached significance ($b = 0.56, SE = 0.30, p = .063, 95\% \text{ CI } [-0.03, 1.16]$). Forgiveness did not predict GASJ ($b = 0.03, SE = 0.08, p = .686, 95\% \text{ CI } [-0.13, 0.19]$) and GASJ did not predict protest behavior ($b = -0.03, SE = 0.02, p = .153, 95\% \text{ CI } [-0.08, 0.01]$). Thus, the specific indirect effect through forgiveness and GASJ was not significant ($b = -.001, SE = .003, 95\% \text{ CI } [-0.011, 0.002]$). The specific indirect effects through forgiveness alone ($b = 0.01, SE = .01, 95\% \text{ CI } [-0.01, 0.05]$) or GASJ alone ($b = 0.01, SE = .01, 95\% \text{ CI } [-0.01, 0.05]$) also did not reach significance. The direct effect was significant ($b = -0.20, SE = .05, p < .001, 95\% \text{ CI } [-0.31, -0.09]$). The pattern for the follow-up was similar.

This time, we could not find a mediation for forgiveness, and consequentially, not for system justification. Besides the small sample size, the reason for this could be that we used different measures than in the previous studies to assess forgiveness and system justification and measured a different outcome, protest behaviors instead of protest intentions. We strongly put our focus on realizing an ecologically valid study design and chose measures and outcomes based on that purpose.

We used a different measure for forgiveness because the old measure might have revealed that the experiment was preprogrammed and that the other participants were not real. Specifically, in this study, we could not ask participants directly whether they forgave the male student because it would have implied that the experiment programmers had anticipated a transgression. This would have threatened the naturalistic setting of the study. Instead, we asked participants more indirectly to indicate what they *felt* at the moment when thinking about the respective other participant and administered, among distractors, “forgiveness”, “leniency” and another German word for forgiveness. Compared to the old items, these items are very abstract. Importantly, they assessed whether participants felt forgiveness, but not whether participants actually forgave, like in the old items. Thus, we found that expressing thanks increased feelings of forgiveness, but “just” feeling forgiveness was not related to system justification or protest behaviors. For system justification, we modified the previous measure for an academic context and had to present it as a separate opinion survey to make it fit in the study. Future research should, therefore, measure whether people actually forgave the transgressor, beyond feeling forgiveness.

Additional References

- Anderson, C., & Berdahl, J. L. (2002). The experience of power: Examining the effects of power on approach and inhibition tendencies. *Journal of Personality and Social Psychology*, 83(6), 1362–1377. <https://doi.org/10.1037/0022-3514.83.6.1362>
- Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén