**Supplementary Materials for**

**Does Stereotype Threat contribute to the Political Knowledge Gender Gap? A Preregistered Replication Study of Ihme and Tausendpfund (2018)**

# Procedure and Measures

The present study abided by the American Political Science Association’s (APSA) Principles and Guidance for Human Subjects Research. The experiment involved no deception and posed no harm to participants. Participants were recruited via a professional survey firm and paid for their participation. Payment was determined by the professional survey firm according to their guidelines. At the beginning of the study, participants read a consent form stating that participation was voluntary, and that data would be collected and stored anonymously. Participants were only able to perform the study after providing consent online. At the end of the study, participants were debriefed on the study aims and research conditions and provided with the contact information of the involved researchers in case of questions.

## 1.1 Field of study/work

Upon providing consent, participants were shown an extensive list of topics and asked to indicate a maximum of four topics which were relevant for their studies and/or work. Participants who selected fields related to Politics (i.e., Politics or History or Public Administration or Government or Sociology) were assigned to the ‘Politics’ group, while participants who did not select any of these five topics were assigned to the ‘Non-Politics’ group. To achieve a similar proportion as in the original study regarding participants in the Politics and Non-Politics groups, the number of participants in each group was controlled by quotas implemented in Qualtrics.

Note that in the preregistration we planned to only include in the Politics group participants who selected “Politics” as one of the topics important for their studies/work. This criterion needed to be relaxed during data collection as we observed a substantial difference in proportion of completed responses in the Politics and Non-Politics groups, making it impossible for the survey company to achieve the required sample size. This modification was preregistered as an amendment to the original preregistration prior to any data analyses (<https://osf.io/kvjeh/?view_only=bcf071c876c5410c85bb0288e4044c9a>, see Appendix A). We note that our final selection of fields related to the group Politics closely resembles the selection of the original study, where participants composing the Politics groups were bachelor students of Political science or Public Administration, or Sociology or Governance.

## 1.2 Political interest

After indicating the important topics for their studies/work, participants answered five items measuring political interest (e.g., “I observe political events with great interest”) on a scale ranging from 1 (not true at all) to 7 (completely true). We used the same scale as in the original study, but the items were presented in English using a translation of the German items proposed by the authors of the scale (Otto and Bacherle 2011). Responses across these five items were averaged to create a composite score of political interest per participant.

## 1.3 Stereotype activation

Immediately before answering the political knowledge test, participants were randomly allocated to one of the three stereotype activation conditions, which were the same way as in the original study. In the control condition (stereotype not activated) there was no mention of gender before the political knowledge test. In the stereotype activated by gender question condition, participants were simply asked to indicate their gender before the political knowledge test. Finally, in the stereotype activated by a gender difference statement condition, the instructions of the political knowledge test were modified to include a statement that the test participants were about to take had shown gender differences in the past (i.e., “Below you will be given a number of questions about your political knowledge, that have been known to produce gender differences with regard to the number of correct answers”). As in the original study, there was no mention regarding the direction of these gender differences.

## 1.4 Gender

Similarly to the original study, participants were asked to report their gender either at the beginning or end of the study depending on which stereotype activation condition they were allocated to. Only participants who identified as male or female were included in the analyses.

## 1.5 Political knowledge scale

The political knowledge scale consisted of 20 questions. As in the original study, all questions were presented with multiple-option responses and one option always allowed participants to indicate they did not know the correct answer. In addition, all questions had a time-limit of 20 seconds, after which the next question was displayed. Questions were presented in a random order and participants were instructed to answer the questions without assistance. Participants scored one point for each question answered correctly.

To devise the political knowledge scale, whenever possible, we tried to use the same questions as in the original study with slight modifications to the American political context. Table S1 below displays the comparison between the items used in the original study and how these items were adapted in the direct replication. Items marked with an asterisk did not have a direct correspondence in the direct replication and were substituted by items from the Annenberg Civics Knowledge Survey (Annenberg Public Policy Center 2019) with slight modifications. Items 4 and 5 were substituted as the direct translation of the item would be more difficult for the American population, considering that the Green Party in Germany is more popular than their equivalent in the United States. Items 9, 15-20 were substituted because they tackled specific political affairs in Germany with no clear correspondence to the American context. A more detailed rationale behind the formulation of the items for the direct replication, including important notes about the context of the items used in the original study, as well as the multiple option choices offered in the direct replication can be found at OSF (<https://osf.io/8feku/?view_only=99a41a96c8cd43c4ab349e44d79919cd>, see Appendix B). The internal consistency of the political knowledge scale was acceptable and similar to the original study (Stage 1: α = .71; Stage 2: α = .7, original study: α = .83).

Table S1

*Political knowledge scale items in the original study and direct replication*

|  |  |  |
| --- | --- | --- |
| **Item** | **Original study (Ihme and Tausendpfund 2018)** | **Direct Replication** |
| 1 | Heiko Maas is member of the following party: | Mike Pompeo is a member of which party: |
| 2 | Andrea Nahles is member of the following party: | Nancy Pelosi is a member of which party: |
| 3 | Gerd Müller is member of the following party: | Wilbur Ross is a member of which party: |
|  4\* | Anton Hofreiter is member of the following party: | Which of the following does not describe one of the three branches of government? |
| 5\* | Katrin Göring-Eckhardt is member of the following party: | Which political party is in control of the U.S. House of Representatives, or are you not sure? |
| 6 | Who is the current Federal minister for family, senior citizens, women and the young? | Who is the current Secretary of Labor? |
| 7 | Who is the current Federal minister for health? | Who is the current Secretary of Health and Human Services? |
| 8 | Who elects the Federal Chancellor of Germany? | Who elects the President of the United States? |
| 9\* | Who commands the so-called ‘Richtlinienkompetenz’? | Which political party is in control of the U.S. Senate, or are you not sure? |
| 10 | Who elects the Federal President of Germany? | Who elects the Vice President of the United States? |
| 11 | During Bundestag elections you have two votes, a first and a second vote.Which of these votes is crucial for the allocation of seats in the Bundestag? | What is the so called “winner takes all” principal in an electoral system? |
| 12 | How high was the percentage of foreign nationals in Germany at the end of 2013? | In 2017, what percentage of the American population was foreign-born? |
| 13 | What is the current (1. quarter 2015) unemployment rate in Germany? | What is the current (1. quarter 2020) unemployment rate in the United States? |
| 14 | Which political department has the highest budget? | Which federal executive department of the United States has the highest budget at their disposal? |
| 15\* | Which one of the following parties argued for the collection and storage of contact data (e.g. telephone, internet) even without concrete cause? | If the president and Supreme Court differ on whether an action by the president is constitutional, who has the final responsibility for determining if the action is constitutional? |
| 16\* | Which one of the following parties argued against raising the top income tax rate? | How much of a majority is required for the U.S. Senate and the House of Representatives to override a presidential veto? |
| 17\* | Which one of the following parties argued for a general speed limit on theautobahn? | If the U.S. Supreme Court rules on a case 5 to 4, does this mean…  |
| 18\* | Which one of the following parties argued against BAföG-payments regardless of parents’ income? | How accurate is the following statement: “the U.S. Supreme Court has held that a citizen has a constitutional right to own a handgun”? |
| 19\* | Which one of the following parties argued for a stricter public control of electricity tariffs? | How accurate is the following statement: “the U.S. Constitution allows a judge to insist that a defendant testify at his own trial”? |
| 20\* | Which one of the following parties argued against a decrease of the statutory retirement age? | How accurate is the following statement: “those who are in the country illegally do not have any rights under the U.S. Constitution”? |

Note: Items of the original study were formulated in German. The translations presented here were done by the original authors and shown in the appendix of the original study. The asterisk denotes items for which a direct correspondence to the American population could not be found. In those cases, the original items were substituted by items from the Annenberg Civics Knowledge Survey with slight adaptations.

## 1.6 Additional questions

Similarly to the original study, after the political knowledge test, participants were asked whether they answered the test without assistance. In addition, to have a more fine grained understanding of how much Politics were important to participants’ studies/work, two control questions were added. Participants were asked to rank the topics they had selected previously in order of importance. Following that, participants were asked to rate the topics they were presented earlier in terms of importance to their studies/work on a scale ranging from 0 (“Not at all important”) to 100 (“Very important”). All responses had to add up together to 100. Participants were then asked to answer demographic questions such as education level, age and gender and were debriefed on the purpose of the study.

# 2. First stage of data collection - detailed results

## 2.1 Sample Stage 1

We recruited 2500 participants at a professional survey company. As is shown in the CONSORT flow diagram below (Figure S1), 1601 responses were terminated early because participants either did not provide consent (*N* = 455), did not commit to answering questions truthfully (*N* = 20), or because quotas for field of study/work had already been completed (*N* = 1126). As in the original study, we excluded participants who indicated they had cheated or used help to answer the political knowledge test (*N* = 84). In addition, as preregistered, we also excluded participants who failed attention check questions (*N* = 143) and who did not identify as male or female (*N* = 1), since the original study only considered differences between these two genders. The final analytical sample after exclusions met the target with 671 participants (*Mage* = 45.11 years, *SD* = 16.97, 47.99% female).

## 2.2 Frequency of participants per conditions

The first stage of data collection reached a distribution of participants per condition similar to the original study. 39.8% of participants were categorized as belonging to the Politics groups, while in the original study 37.4% of participants belonged to the Politics group. In addition, 48% of our sample was composed by female participants, while in the original study 58.9% of participants were female. Table S2 shows the frequency of participants per gender, stereotype activation condition and field of study/work. The relative proportion of participants per condition is shown between parentheses. Numbers in bold reflect the relative distribution of participants across conditions in the original study.

Table S2

*Frequency of participants per gender, stereotype activation condition and field of study/work (Stage 1)*

|  |
| --- |
| Condition |
|  | Stereotype not activated | Stereotype activated by gender question | Stereotype activated by gender statement |
|  |   | Politics |   |
| **Female** | N=31 (4.62%)  | N=38 (5.66%) | N=32 (4.77%) |
|  | **(5.57%)** | **(6.37%)** | **(4.78%)** |
| **Male** | N=54 (8.05%) | N=44 (6.56%) | N=68 (10.13%) |
|  | **(8.22%)** | **(4.78%)** | **(7.7%)** |
|  |   | Non-Politics |   |
| **Female** | N=67 (9.98%) | N=77 (11.47%) | N=77 (11.47%) |
|  | **(13.79%)** | **(14.59%)** | **(13.79%)** |
| **Male** | N=59 (8.8%) | N=65 (9.69%) | N=59 (8.8%) |
|  | **(5.83%)** | **(5.83%)** | **(8.75%)** |

## 2.3 Statistical analyses

Similar to the original study, we conducted a 2 (gender: male vs female) × 2 (field of work/study: Political vs Non-Political) × 3 (Gender Stereotype Activation: stereotype not activated vs. stereotype activated by gender-question vs. stereotype activated by gender difference statement) ANOVA with participants’ score in the political knowledge test as the dependent variable. Results are shown in table S3. No significant interaction between stereotype activation and gender was found, suggesting the effect of stereotype activation on the gender gap in political knowledge was not successfully replicated.

Table S3

*Fixed-Effects ANOVA results using political knowledge as the dependent variable (Stage 1)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 3424.49 | 1 | 3424.49 | 239.02 | < .001 |   |   |
| Gender | 85.14 | 1 | 85.14 | 5.94 | .015 | .01 | [.00, .02] |
| Field of study/work | 0.06 | 1 | 0.06 | 0.00 | .949 | .00 | [.00, 1.00] |
| Stereotype activation | 5.02 | 2 | 2.51 | 0.18 | .839 | .00 | [.00, .00] |
| Gender x Field of study/work | 60.56 | 1 | 60.56 | 4.23 | .040 | .01 | [.00, .02] |
| Gender x Stereotype activation | 9.58 | 2 | 4.79 | 0.33 | .716 | .00 | [.00, .01] |
| Field of study/work x Stereotype activation | 66.82 | 2 | 33.41 | 2.33 | .098 | .01 | [.00, .02] |
| Gender x Field of study/work x Stereotype activation | 58.88 | 2 | 29.44 | 2.05 | .129 | .01 | [.00, .02] |
| Error | 9441.79 | 659 | 14.33 |   |   |   |   |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively.

The replication was still not successful when we included political interest as a covariate (see Table S4). We note, however, that the interaction between gender and field of study/work was significant. Post-hoc pairwise comparisons using the *emmeans* package in R (Lenth et al. 2018) revealed that males in the Politics group performed significantly better in the political knowledge test than males in the Non-Politics group, *t*(658) = 3.13, *p* = .002, whereas no differences were observed between female participants in these two groups, *t*(658) = 1.57, *p* = .11. We also observed a significant interaction between field of study/work and stereotype activation. Post-hoc pairwise comparisons showed that participants in the Politics group performed better in the political knowledge test than participants in the Non-Politics group in both the stereotype activated by gender question condition, *t*(658) = 2.5, *p* = .01, and gender statement condition, *t*(658) = 2.4, *p* = .01, while there were no differences between participants from different fields of study/work in the control condition, *t*(658) = .79, *p* = .43. Finally, the three-way interaction between gender, field of study/work and stereotype activation was also significant. Post-hoc pairwise comparisons controlling for multiple comparisons with the Bonferroni method showed that in the Non-Politics group, male participants scored higher in the political knowledge test than female participants in the stereotype activated by gender statement condition, *t*(658) = 3.2, *p* = .034, but no gender differences were observed in the control condition, *t*(658) = 1.52, *p* = 1, or in the stereotype activated by gender question condition, *t*(658) = 2.87, *p* = .10. In contrast, in the Politics group, males performed better than females in the control condition, *t*(658) = 4.39, *p* < .001, but no gender differences emerged in the stereotype activated by gender question condition, *t*(658) = 1.03, *p* = 1, or in the stereotype activated by gender statement condition, *t*(658) = 2.55, *p* = .27.

Table S4

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable and political interest as a covariate (Stage 1)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 521.80 | 1 | 521.80 | 42.11 | < .001 |   |   |
| Political interest | 1288.20 | 1 | 1288.20 | 103.96 | < .001 | .14 | [.10, .18] |
| Gender | 28.62 | 1 | 28.62 | 2.31 | .129 | .00 | [.00, .01] |
| Field of study/work | 15.38 | 1 | 15.38 | 1.24 | .266 | .00 | [.00, .01] |
| Stereotype activation | 15.99 | 2 | 8.00 | 0.65 | .525 | .00 | [.00, .01] |
| Gender x Field of study/work | 77.02 | 1 | 77.02 | 6.22 | .013 | .01 | [.00, .03] |
| Gender x Stereotype activation | 17.12 | 2 | 8.56 | 0.69 | .501 | .00 | [.00, .01] |
| Field of study/work x Stereotype activation | 81.07 | 2 | 40.53 | 3.27 | .039 | .01 | [.00, .02] |
| Gender x Field of study/work x Stereotype activation | 78.42 | 2 | 39.21 | 3.16 | .043 | .01 | [.00, .02] |
| Error | 8153.59 | 658 | 12.39 |   |   |   |   |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively.

# 3. Detailed results for the pooled data

## 3.1 Sample Stage 2

For the second stage of data collection, 1625 additional participants were recruited at the same professional survey used in the first stage. As is shown in the CONSORT flow diagram below (Figure S2), 218 responses were terminated earlier because participants did not provide consent (*N* = 125), did not commit to answering the study truthfully (*N* = 13), or because quotas were already full (*N* = 80). 293 participants were excluded because they did not complete the study. Following the preregistration, 283 participants were excluded for either cheating on the political knowledge test (*N* = 105), failing attention check questions (*N* = 174), not identifying as male/female (*N* = 3), or being younger than 18 years old (*N* = 1). The final analytical sample after exclusions met the target with 831 participants (*Mage* = 46.49 years, *SDage* = 17.63, 49.34% female).

## 3.2 Frequency of participants per conditions

When we pooled the data from both the first and second stage of data collection (N=1502), the distribution of participants per conditions was still similar to the distribution of the original study. 39.75% of participants were assigned to the Politics group and 48.74% of participants were female. Table S5 shows the frequency of participants per gender, stereotype activation condition and field of study/work. The relative proportion of participants per condition is shown between parentheses. Numbers in bold reflect the relative distribution of participants across conditions in the original study.

Table S5

*Frequency of participants per gender, stereotype activation condition and field of study/work (pooled sample)*

|  |  |
| --- | --- |
|  | Condition |
|  | Stereotype not activated | Stereotype activated by gender question | Stereotype activated by gender statement |
|  |   | Politics |   |
| **Female** | N=68 (4.53%) | N=78 (5.19%) | N=80 (5.33%) |
|  | **(5.57%)** | **(6.37%)** | **(4.78%)** |
| **Male** | N=118 (7.86%) | N=125 (8.32%) | N=128 (8.52%) |
|  | **(8.22%)** | **(4.78%)** | **(7.7%)** |
|  |   | Non-Politics |   |
| **Female** | N=166 (11.05%) | N=161 (10.72%) | N=179 (11.92%) |
|  | **(13.79%)** | **(14.59%)** | **(13.79%)** |
| **Male** | N=144 (9.59%) | N=127 (8.45%) | N=128 (8.52%) |
|  | **(5.83%)** | **(5.83%)** | **(8.75%)** |

##

## 3.3 Statistical analyses

We conducted again a 2 (gender: male vs female) × 2 (field of work/study: Political vs Non-Political) × 3 (Gender Stereotype Activation: stereotype not activated vs. stereotype activated by gender-question vs. stereotype activated by gender difference statement) ANOVA with participants’ score in the political knowledge test as the dependent variable. Results are shown in table S6. In line with the results obtained at the first stage of data collection, we could not replicate the effect of gender stereotype threat on gender differences in political knowledge.

Table S6

*Fixed-Effects ANOVA results using political knowledge as the dependent variable (pooled sample)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 7719.42 | 1 | 7719.42 | 574.20 | < .001 |   |   |
| Gender | 447.82 | 1 | 447.82 | 33.31 | < .001 | .02 | [.01, .04] |
| Field of study/work | 27.44 | 1 | 27.44 | 2.04 | .153 | .00 | [.00, .01] |
| Stereotype activation | 2.82 | 2 | 1.41 | 0.10 | .900 | .00 | [.00, .00] |
| Gender x Field of study/work | 16.44 | 1 | 16.44 | 1.22 | .269 | .00 | [.00, .01] |
| Gender x Stereotype activation | 31.00 | 2 | 15.50 | 1.15 | .316 | .00 | [.00, .01] |
| Field of study/work x Stereotype activation | 32.13 | 2 | 16.07 | 1.19 | .303 | .00 | [.00, .01] |
| Gender x Field of study/work x Stereotype activation | 8.51 | 2 | 4.25 | 0.32 | .729 | .00 | [.00, .00] |
| Error | 20031.08 | 1490 | 13.44 |   |   |   |   |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively.

Table S7 displays a comparison between the replication results and the results from the original study. Results were obtained with a 2 (gender: male vs female) × 2 (field of work/study: Political vs Non-Political) × 3 (Gender Stereotype Activation: stereotype not activated vs. stereotype activated by gender-question vs. stereotype activated by gender difference statement) ANCOVA with participants’ score in the political knowledge test as the dependent variable and participants’ political interest as a covariate.

We found no evidence that gender stereotype activation negatively influenced women’s performance in the political knowledge test when conducting Bonferroni-adjusted pairwise comparisons using the *emmeans* package in R. Female participants did not have higher scores in the political knowledge test in the stereotype not activated condition than in the stereotype activated by gender difference statement condition, *t*(1489) = -1.799, *p* = .072, *padjusted* = .217, or the stereotype activated by gender question condition, *t*(1489) = -1.895, *p* = .058, *padjusted* = .175. Likewise, gender stereotype activation did not boost male’s performance in the political knowledge test. Male participants’ scores in the political knowledge test were not significantly lower in the stereotype not activated condition than in the stereotype activated by gender difference statement condition, *t*(1489) = -.264, *p* = .792, *padjusted* = 1, or than in the stereotype activated by gender question condition, *t*(1489) = 1.242, *p* = .215, *padjusted* = .644.[[1]](#footnote-1)

Table S7.

*Fixed-Effects ANCOVA results in the original study and in the direct replication.*

|  |  |  |
| --- | --- | --- |
|  | **Original study (Ihme and Tausendpfund 2018)** | **Direct replication** |
| Predictor | Sum of Squares (Type III) | *F* | *p* | partial η2 | Sum of Squares (Type III) | *F* | *p* | partial η2 |
| (Intercept) | 359.3 | 32.7 | < .001 |  | 1157.9 | 97.33 | < .001 |  |
| Political Interest | 1119.23 | 101.86 | < .001 | .219 | 2317.15 | 194.78 | < .001 | .12 |
| Gender | 211.33 | 19.2 | < .001 | .05 | 340.33 | 28.61 | < .001 | .02 |
| Field of study/work | 270.02 | 24.57 | < .001 | .063 | 1.5 | .12 | .72 | .00 |
| Stereotype activation | 5.11 | .23 | .79 | .001 | 6.3 | .27 | .77 | .00 |
| Gender x Field of study/work | 7.4 | .67 | .41 | .002 | 12.51 | 1.05 | .3 | .00 |
| Gender x Stereotype activation | 135.59 | 6.17 | .002 | .033 | 28.92 | 1.22 | .3 | .00 |
| Field of study/work x Stereotype activation | 79.08 | 3.6 | .028 | .019 | 26.82 | 1.13 | .32 | .00 |
| Gender x Field of study/work x Stereotype activation | 6.14 | .28 | .76 | .002 | 5.15 | .22 | .81 | .00 |
| Error | 3999.726 |  |  |  | 17713.93 |  |  |  |
| N | 377 |  |  |  | 1502 |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. Results from the original study (Ihme & Tausendpfund, 2018) are obtained from the original paper’s [supplementary materials](https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/gender-differences-in-political-knowledge-bringing-situation-back-in/4897AD1BEF5BC95E164F08BDFB5ED1AF#supplementary-materials). |

# 4. Exploratory analyses

As our results did not replicate the effect of stereotype activation on the gender gap in political knowledge, we conducted several exploratory analyses.

## 4.1 Analyses of “don’t know” and incorrect answers in the political knowledge test

First, as in the original study, we tested whether gender stereotype activation could have influenced participants’ answering behavior in the political knowledge test. According to the original authors, stereotype threat could lead female participants to engage in an avoidance-oriented strategy, which ultimately could impair their performance in the political knowledge test. This would be reflected for example on the number of questions in the political knowledge test in which they used the “don’t know” option to respond or in the number of questions they did not attempt to answer.

Following the original study, we conducted a 2 (gender: female, male) x 3 (Gender Stereotype Activation: stereotype not activated vs. stereotype activated by gender-question vs. stereotype activated by gender difference statement) ANCOVA where the dependent variable was the number of questions in which participants answered with the “don’t know” option and political interest was a covariate. Tables S8 and S9 below display the results. Contrasting with the original study, we find that the main effect of gender on the number of “don’t know” options used was significant, *F*(1, 1495) = 33.17, *p* < .001, partial η² = .02, 95% CI = [.01, .04], with female participants using this option on average more often than their male counterparts (*Mfemale* = 5.78, *SD* = 4.75, *Mmale* = 3.53, *SD* = 3.76). However, the interaction between stereotype activation and gender was not significant, *F*(2, 1495) = 2.65, *p* = .071, partial η² = .004, 95% CI = [.00, .01], suggesting gender stereotype activation did not lead to a difference in number of “don’t know” answers by males and females.

Table S8.

*Fixed-Effects ANCOVA results using frequency of “don’t know” responses as the dependent variable and political interest as a covariate*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 14375.9 | 1 | 14375.9 | 1010.53 | < .001 |   |   |
| Political interest | 5979.8 | 1 | 5979.8 | 420.34 | < .001 | .22 | [.19, .25] |
| Gender | 471.8 | 1 | 471.8 | 33.17 | < .001 | .02 | [.01, .04] |
| Stereotype activation | 32.42 | 2 | 16.21 | 1.14 | .32 | .002 | [.00, .00] |
| Gender x Stereotype activation | 75.5 | 2 | 37.76 | 2.65 | .071 | .004 | [.00, .01] |
| Error | 21268.1 | 1495 | 14.23 |   |   |   |   |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. |

Table S9.

*Means and standard deviations of the number of “don’t know” responses in the political knowledge test by gender and stereotype activation condition*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Stereotype not activated | Stereotype activated by gender question | Stereotype activated by gender difference statement |
|  | Mean (SD) | *N* | Mean (SD) | *N* | Mean (SD) | *N* |
| Female | 6.07 (4.85) | 234 | 5.68 (4.78) |  239 | 5.62 (4.63) | 259 |
| Male | 3.42 (3.78) | 262 | 4.02 (3.99) |  252 |  3.16 (3.46) | 256 |

We repeated these analyses, using the number of incorrect answers as a dependent variable. Table S10 and S11 below display the results. We find a significant main effect of gender on the number of incorrect answers, *F*(1, 1495) = 26.15, *p* < .001, partial η² = .02, 95% CI = [.01, .03], with male participants getting a higher number of incorrect answers than females (*Mfemale* = 5.21, *SD* = 2.40, *Mmale* = 6.21, *SD* = 2.25). Again, no significant interaction term between stereotype activation and gender was found, *F*(2, 1495) = 1.10, *p* = .33, partial η² = .00, 95% CI = [.00, .01], suggesting gender stereotype activation cannot account for the difference in number of incorrect answers between males and females.

Table S10.

*Fixed-Effects ANCOVA results using frequency of incorrect responses as the dependent variable and political interest as a covariate*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 1155.94 | 1 | 1155.94 | 233.32 | < .001 |  |  |
| Political interest | 675.66 | 1 | 675.66 | 136.38 | < .001 | .08 | [.06, .11] |
| Gender | 129.54 | 1 | 129.54 | 26.15 | < .001 | .02 | [.01, .03] |
| Stereotype activation | 9.18 | 2 | 4.59 | 0.93 | .396 | .00 | [.00, .01] |
| Gender x Stereotype activation | 10.92 | 2 | 5.46 | 1.10 | .332 | .00 | [.00, .01] |
| Error | 7406.65 | 1495 | 4.95 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. |

Table S11.

*Means and standard deviations of the number of incorrect responses in the political knowledge test by gender and stereotype activation condition*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Stereotype not activated | Stereotype activated by gender question | Stereotype activated by gender difference statement |
|  | Mean (SD) | *N* | Mean (SD) | *N* | Mean (SD) | *N* |
| Female | 5.06 (2.54) | 234 | 5.20 (2.39) |  239 | 5.36 (2.29) | 259 |
| Male | 6.32 (2.54) | 262 | 6.09 (2.24) |  252 |  6.22 (2.26) | 256 |

Given that “don’t know” and incorrect answers differ on a conceptual and empirical level, we may assume that they do not equally represent simple absence of knowledge (Mondak, 1999). In line with Mondak and Anderson (2004), we find a relative reluctance of men to admit that they “don’t know” and a relative propensity of men to answer incorrectly. However, given that Ihme and Tausendpfund (2008) had not found an effect of gender on the amount of “don’t know” answers, we cannot rule out that national differences on answering patterns between Germany (original study) and the U.S. (replication study) could have influenced the results. To control for these potential differences, we included the proportion of “don’t know” answers relative to incorrect answers (sum of “don’t know” answers / sum of “don’t know” answers + sum of incorrect answers) as a covariate in the main analysis. Results of a 2 (gender) × 2 (field of work/study) × 3 (Gender Stereotype Activation) ANCOVA with participants’ score in the political knowledge test as the dependent variable and controlling for the proportion of “don’t know” to incorrect answers as well as participants’ political interest are displayed in Table S12. The interaction between gender and stereotype activation was still insignificant *F*(2, 1486) = 0.24, *p* = .751, partial η² = .00, 95% CI = [.00, .00]. Thus, our results are robust even when controlling for potential differences in the use of “don’t know” answers between Germany and the U.S.

Table S12

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable, proportion of “don’t knows” to incorrect answers and political interest as a covariate*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 4957.05 | 1 | 4957.05 | 579.91 | < .001 |  |  |
| Political interest | 231.31 | 1 | 231.31 | 27.06 | < .001 | .02 | [.01, .03] |
| Proportion “don’t know”  | 4832.62 | 1 | 4832.62 | 565.35 | < .001 | .28 | [.24, .31] |
| Gender | 107.94 | 1 | 107.94 | 12.63 | < .001 | .01 | [.00, .02] |
| Field of study/work | 0.06 | 1 | 0.06 | 0.01 | .934 | .00 | [.00, .00] |
| Stereotype activation | 0.55 | 2 | 0.28 | 0.03 | .968 | .00 | [.00, .00] |
| Gender x Field of study/work | 2.03 | 1 | 2.03 | 0.24 | .626 | .00 | [.00, .00] |
| Gender x Stereotype activation | 4.90 | 2 | 2.45 | 0.29 | .751 | .00 | [.00, .00] |
| Field of study/work x Stereotype activation | 6.09 | 2 | 3.04 | 0.36 | .700 | .00 | [.00, .00] |
| Gender x Field of study/work x Stereotype activation | 0.90 | 2 | 0.45 | 0.05 | .949 | .00 | [.00, .00] |
| Error | 12702.24 | 1486 | 8.55 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. |

## 4.2 Analyses of attempted answers in the political knowledge test

We also conducted analyses using the number of questions participants attempted to answer as the dependent variable and controlling for political interest. Table S13 and S14 below display the results. Our findings are again at odds with the original study, as we find a significant main effect of gender on the number of attempted questions in the political knowledge test, *F*(1, 1495) = 27.43, *p* < .001, partial η² = .02, 95% CI = [.01, .03], but no significant interaction between gender and stereotype activation, *F*(2, 1495) = 25.59, *p* = .14, partial η² = .003, 95% CI = [.00, .01]. Our results suggest that males in general attempted to answer more questions than females (*Mfemale* = 12.89, *SD* = 4.73, *Mmale* = 15.1, *SD* = 4.01), but there was no difference in answering behavior between genders across the stereotype activation conditions.

Table S13.

*Fixed-Effects ANCOVA results using frequency of attempted responses in the political knowledge test as the dependent variable and political interest as a covariate*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 5731.3 | 1 | 5731.3 | 378.27 | < .001 |  |   |
| Political interest | 5978.31 | 1 | 5978.31 | 394.57 | < .001 | .209 | [.18, .24] |
| Gender | 415.64 | 1 | 415.64 | 27.43 | < .001 | .018 | [.01, .03] |
| Stereotype activation | 25.6 | 2 | 12.8 | .84 | .43 | .001 | [.00, .00] |
| Gender x Stereotype activation | 59.18 | 2 | 29.59 | 1.95 | .14 | .003 | [.00, .01] |
| Error | 22651.14 | 1495 | 15.15 |   |   |   |   |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. |

Table S14.

*Means and standard deviations of the number of attempted questions in the political knowledge test by gender and stereotype activation condition*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Stereotype not activated | Stereotype activated by gender question | Stereotype activated by gender difference statement |
|  | Mean (SD) | *N* | Mean (SD) | *N* | Mean (SD) | *N* |
| Female | 12.63 (4.89) | 234 | 12.95 (4.68) |  239 | 13.06 (4.63) | 259 |
| Male | 15.16 (4.08) | 262 | 14.64 (4.15) |  252 |  15.47 (3.77) | 256 |

## 4.3 Political interest

As in the original paper, we also tested where there were any differences regarding political interest between the genders and the gender stereotype activation conditions. We ran a 2 (gender: female, male) x 3 (Gender Stereotype Activation: stereotype not activated vs. stereotype activated by gender-question vs. stereotype activated by gender difference statement) ANOVA with political interest as the dependent variable. Results can be seen in Table S15 and closely resemble those reported in the original study. First, we found a main effect of gender indicating that male participants are generally more interested in politics than female participants (*Mfemale* = 1.42, *SD* = 1.72; *Mmale* = 4.95, *SD* = 1.56). In addition, we also found a main effect of field of study/work, indicating that those who were categorized in the Politics group are generally more interested in the topic than those participants categorized in the non-Politics group (*MPolitics* = 5.23, *SD* = 1.49; *MNonPolitics* = 4.33, *SD* = 1.67). In contrast with the original study, however, the interaction between gender and field of study/ work is not significant. All the other interactions were not significant, in line with the results of the original study. In addition, as reported in the original paper, political interest significantly correlates with political knowledge (*r* = .4, *p* < .001), suggesting the more individuals feel interested about political topics, the better their performance in the political knowledge test.

Table S15.

*Fixed-Effects ANOVA results using political interest as the dependent variable*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 2943.4 | 1 | 2943.4 | 1160.27 | <.001 |  |   |
| Gender | 11.76 | 1 | 11.76 | 4.64 | .031 | .003 | [.00, .01] |
| Field of study/work | 26.36 | 1 | 26.36 | 10.39 | .001 | .007 | [.00, .02] |
| Stereotype activation | 1.83 | 2 | .91 | .36 | .7 | .000 | [.00, .00] |
| Gender x Field  | .44 | 1 | .44 | .17 | .68 | .000 | [.00, .00] |
| Gender x Stereotype activation | .14 | 2 | .07 | .028 | .97 | .000 | [.00, .00] |
| Field x Stereotype activation | .77 | 2 | .39 | .15 | .86 | .000 | [.00, .00] |
| Gender x Field x Stereotype activation | 1.28 | 2 | .64 | .25 | .78 | .000 | [.00, .00] |
| Error | 3779.84 | 1490 | 2.54 |   |   |   |   |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. |

## 4.4. Recoding the Politics group

We also explored whether the few differences between our direct replication and the original study with regards to study design could have led to the discrepancy in results. The Politics group in our direct replication included one additional discipline compared to the original study (i.e., History). We recoded the Politics group to be more similar to the original study and only include the following disciplines: Politics, Public Administration, Government and Sociology. However, we were still not able to replicate the original findings. Results of a 2 (gender) × 2 (field of work/study) × 3 (Gender Stereotype Activation) ANCOVA controlling for political interest revealed a non-significant interaction between gender and stereotype activation *F*(2, 1489) = 1.34, *p* = .26, partial η² = .002, 95% CI = [.00, .01]. Full results are displayed in Table S16 below.

Table S16.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable and political interest as a covariate (Politics groups recoded)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 1166 | 1 | 1166 | 98.36 | < .001 |   |   |
| Political interest | 2181.9 | 1 | 2181.9 | 184.06 | < .001 | .11 | [.09, .14] |
| Gender | 431.9 | 1 | 431.9 | 36.44 | < .001 | .024 | [.01, .04] |
| Field of study/work | 26.81 | 1 | 26.81 | 2.26 | .13 | .002 | [.00, .01] |
| Stereotype activation | 13.2 | 2 | 6.6 | .56 | .57 | .001 | [.00, .00] |
| Gender x Field of study/work | 0.19 | 1 | .19 | .016 | .9 | .000 | [.00, .00] |
| Gender x Stereotype activation | 31.7 | 2 | 15.8 | 1.34 | .26 | .002 | [.00, .01] |
| Field of study/work x Stereotype activation | 7.1 | 2 | 3.5 | .3 | .74 | .000 | [.00, .00] |
| Gender x Field of study/work x Stereotype activation | 0.92 | 2 | .46 | .039 | .962 | .000 | [.00, .00] |
| Error | 17650.8 | 1489 | 11.85 |   |   |   |   |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. |

## 4.5 Removing field of study/work from analysis

Next, we ran the same analyses excluding field of study/work as a factor, as this variable was not directly important for the present replication. Results remained the same and the interaction between gender and stereotype activation was still insignificant *F*(2, 1495) = 1.99, *p* = .14, partial η² = .003, 95% CI = [.00, .01]. Full results are displayed in Table S17 below.

Table S17.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable and political interest as a covariate (Field of study/work not included)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 1162.8 | 1 | 1162.8 | 96.27 | <.001 |   |   |
| Political interest | 2932.62 | 1 | 2932.62 | 242.8 | <.001 | .14 | [.11, .17] |
| Gender | 702.15 | 1 | 702.15 | 58.13 | <.001 | .037 | [.02, .05] |
| Stereotype activation | 37.65 | 2 | 18.83 | 1.56 | .21 | .002 | [.00, .01] |
| Gender x Stereotype activation | 48.21 | 2 | 24.1 | 1.99 | .14 | .003 | [.00, .01] |
| Error | 18058.55 | 1495 | 12.08 |   |   |   |   |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. |

## 4.6 Participants’ ranking of Politics as an important topic for their studies/work

We note that controlling for how important participants thought Politics was to their studies/work did not change the results. Results of a 2 (gender) × 2 (field of work/study) × 3 (Gender Stereotype Activation) ANCOVA with participants’ score in the political knowledge test as the dependent variable and controlling for participants’ political interest as well as participants’ ranking of Politics as an important topic for their studies/work are displayed in Table S18. The interaction between gender and stereotype activation was still insignificant *F*(2, 1488) = 1.22, *p* = .3, partial η² = .00, 95% CI = [.00, .01].

Table S18.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable, political interest and ranking of Politics as covariates*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | df | MeanSquare | F | p | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 1184.59 | 1 | 1184.59 | 99.66 | <.001 |   |   |
| Political interest | 2133.38 | 1 | 2133.38 | 179.48 | <.001 | .11 | [.08, .14] |
| Ranking Politics  | 27.22 | 1 | 27.22 | 2.29 | .130 | .00 | [.00, .01] |
| Gender | 342.42 | 1 | 342.42 | 28.81 | <.001 | .02 | [.01, .03] |
| Field of study/work | 0.07 | 1 | 0.07 | 0.01 | .941 | .00 | [.00, .00] |
| Stereotype activation | 6.24 | 2 | 3.12 | 0.26 | .769 | .00 | [.00, .00] |
| Gender x Field  | 10.22 | 1 | 10.22 | 0.86 | .354 | .00 | [.00, .01] |
| Gender x Stereotype activation | 28.96 | 2 | 14.48 | 1.22 | .296 | .00 | [.00, .01] |
| Field x Stereotype activation | 26.50 | 2 | 13.25 | 1.11 | .328 | .00 | [.00, .01] |
| Gender x Field x Stereotype activation | 5.02 | 2 | 2.51 | 0.21 | .810 | .00 | [.00, .00] |
| Error | 17686.71 | 1488 | 11.89 |   |   |   |   |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. |

## 4.7 Participants’ attributed importance to Politics

Table S19 displays the 2 (gender) × 2 (field of work/study) × 3 (Gender Stereotype Activation) ANCOVA with participants’ score in the political knowledge test as the dependent variable and controlling for participants’ political interest as well as participants’ attributed importance to Politics. Similar to the results reported above, the interaction between gender and stereotype activation was still insignificant if we included participants’ attributed importance to Politics as an additional covariate (*F*(2, 1488) = 1.38, *p* = .26, partial η² = .00, 95% CI = [.00, .01].

Table S19.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable, political interest and importance attributed to Politics as covariates*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | df | MeanSquare | F | p | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 1149.06 | 1 | 1149.06 | 98.85 | <.001 |   |   |
| Political interest | 2666.13 | 1 | 2666.13 | 229.37 | <.001 | .13 | [.10, .17] |
| Importance Of Politics | 417.65 | 1 | 417.65 | 35.93 | <.001 | .02 | [.01, .04] |
| Gender | 327.70 | 1 | 327.70 | 28.19 | <.001 | .02 | [.01, .03] |
| Field of study/work | 3.16 | 1 | 3.16 | 0.27 | .602 | .00 | [.00, .00] |
| Stereotype activation | 9.43 | 2 | 4.71 | 0.41 | .667 | .00 | [.00, .00] |
| Gender x Field | 17.13 | 1 | 17.13 | 1.47 | .225 | .00 | [.00, .01] |
| Gender x Stereotype activation | 32.08 | 2 | 16.04 | 1.38 | .252 | .00 | [.00, .01] |
| Field x Stereotype activation | 30.99 | 2 | 15.49 | 1.33 | .264 | .00 | [.00, .01] |
| Gender x Field x Stereotype activation | 5.71 | 2 | 2.85 | 0.25 | .782 | .00 | [.00, .00] |
| Error | 17296.28 | 1488 | 11.62 |   |   |   |   |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively.

## 4.8 Sample composition

Another factor that differed between the present replication and the original study was the sample composition. Ihme and Tausendpfund’s (2018) sample consisted mainly of university students whereas we collected data from a more heterogeneous national sample. To rule out potential bias, we composed a new sample (*N* = 522) by excluding participants who were older than 43 years old (i.e., the median age value in our sample) or whose education attainment was equal to a high school level or lower. Thus, this new sample was composed of young and highly educated participants and resembled the original sample with regard to educational level, age (original sample: *Mage* = 31, *SDage* = 12.04; replication subsample: *Mage* = 31.26, *SDage* = 7.12) and gender (original sample:58.9% female; replication subsample: 51.92% female). Yet, results of a 2 (gender) × 2 (field of work/study) × 3 (Gender Stereotype Activation) ANCOVA controlling for political interest revealed a non-significant interaction between gender and stereotype activation *F*(2, 509) = 0.15, *p* = .86, partial η² = .001, 95% CI = [.00, 1], suggesting an unsuccessful replication of the effect of gender stereotype activation on political knowledge even with a sample more comparable to that of the original study. Full results are displayed in Table S20 below.

Table S20.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable, political interest and importance attributed to Politics as covariates on a young and educated subsample*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | df | MeanSquare | F | p | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 459.76 | 1 | 459.76 | 42.52 | <.001 |  |  |
| Political interest | 394.31 | 1 | 394.31 | 36.47 | <.001 | .07 | [.03, .11] |
| Gender | 21.28 | 1 | 21.28 | 1.97 | .161 | .00 | [.00, .02] |
| Field of study/work | 11.52 | 1 | 11.52 | 1.07 | .302 | .00 | [.00, .02] |
| Stereotype activation | 3.99 | 2 | 2.00 | 0.18 | .832 | .00 | [.00, .01] |
| Gender x Field | 0.11 | 1 | 0.11 | 0.01 | .918 | .00 | [.00, .00] |
| Gender x Stereotype activation | 2.56 | 2 | 1.28 | 0.12 | .888 | .00 | [.00, .01] |
| Field x Stereotype activation | 1.88 | 2 | 0.94 | 0.09 | .917 | .00 | [.00, .01] |
| Gender x Field x Stereotype activation | 3.25 | 2 | 1.62 | 0.15 | .860 | .00 | [.00, .01] |
| Error | 5503.10 | 509 | 10.81 |  |  |  |  |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively.

## 4.9 Measurement Invariance

We also conducted invariance tests for the political knowledge scale for gender using Item Response Theory’s (IRT) 2-PL model. To assess invariance, we used the non-parametric Lord’s Chi-Square-Test χ2 because it allows for non-uniform (crossing) differential item functioning. Results suggest that only 2 items are invariant by gender (item 9 and 11) even when controlling for multiple comparisons (see Table S21). Females score -1.125 points lower, on average, than males but Figure S1 shows the instruments’ systematic bias is contingent on the participants’ knowledge level (θ) such that non-invariance is stronger at average scores and weaker at the extremes of the political knowledge continuum. This means that the instrument systematically favors males over female respondents with average political knowledge whereas this bias is attenuated as respondents show lower or higher political knowledge.

Table S21

*Lord’s Chi-Square-Test χ2*

|  |  |  |  |
| --- | --- | --- | --- |
| Item | *Lord’s χ2* | P-value |  |
| Item 1 | 3399.683 | 0 | \*\*\* |
| Item 2 | 3806.288 | 0 | \*\*\* |
| Item 3 | 206.3704 | 0 | \*\*\* |
| Item 4 | 23.6708 | 0 | \*\*\* |
| Item 5 | 1489.153 | 0 | \*\*\* |
| Item 6 | 3977.556 | 0 | \*\*\* |
| Item 7 | 1298.857 | 0 | \*\*\* |
| Item 8 | 17.7129 | 0.0001 | \*\*\* |
| Item 9 | 0.7833 | 0.6759 |  |
| Item 10 | 597.9978 | 0 | \*\*\* |
| Item 11 | 3.6744 | 0.1593 |  |
| Item 12 | 2695.477 | 0 | \*\*\* |
| Item 13 | 3781.547 | 0 | \*\*\* |
| Item 14 | 4554.345 | 0 | \*\*\* |
| Item 15 | 2004.235 | 0 | \*\*\* |
| Item 16 | 2131.431 | 0 | \*\*\* |
| Item 17 | 3768.225 | 0 | \*\*\* |
| Item 18 | 1291.625 | 0 | \*\*\* |
| Item 19 | 2219.05 | 0 | \*\*\* |
| Item 20 | 551.6933 | 0 | \*\*\* |

Figure S1. Measurement invariance of political knowledge scale with Item Response Theory.



## 4.10 Analyses of variance with mean-centered covariates and orthogonal contrasts

All analyses of variance previously reported were repeated with orthogonal contrasts for factor variables and mean-centered covariates whenever covariates were included in the respective model. Results did not change substantially and the focal interaction term was still not significant. See Tables S22 – S35 for all results.

Table S22.

*Fixed-Effects ANOVA results using political knowledge as the dependent variable (Stage 1)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 46588.36 | 1 | 46588.36 | 3251.68 | < .001 |  |  |
| Gender1 | 729.39 | 1 | 729.39 | 50.91 | < .001 | .07 | [.04, .11] |
| Field of study/work1 | 381.02 | 1 | 381.02 | 26.59 | < .001 | .04 | [.02, .07] |
| Stereotype activation1 | 35.50 | 2 | 17.75 | 1.24 | .290 | .00 | [.00, .02] |
| Gender x Field of study/work | 9.13 | 1 | 9.13 | 0.64 | .425 | .00 | [.00, .01] |
| Gender x Stereotype activation | 43.83 | 2 | 21.91 | 1.53 | .217 | .00 | [.00, .02] |
| Field of study/work x Stereotype activation | 20.19 | 2 | 10.10 | 0.70 | .495 | .00 | [.00, .01] |
| Gender x Field of study/work x Stereotype activation | 58.88 | 2 | 29.44 | 2.05 | .129 | .01 | [.00, .02] |
| Error | 9441.79 | 659 | 14.33 |  |  |  |  |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Factors were coded with orthogonal contrasts.

Table S23.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable and political interest as a covariate (Stage 1)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 45876.63 | 1 | 45876.63 | 3702.27 | < .001 |  |  |
| Political interest1 | 1288.20 | 1 | 1288.20 | 103.96 | < .001 | .14 | [.09, .18] |
| Gender2 | 488.42 | 1 | 488.42 | 39.42 | < .001 | .06 | [.03, .09] |
| Field of study/work2 | 129.10 | 1 | 129.10 | 10.42 | .001 | .02 | [.00, .04] |
| Stereotype activation2 | 41.61 | 2 | 20.80 | 1.68 | .187 | .01 | [.00, .02] |
| Gender x Field of study/work | 10.78 | 1 | 10.78 | 0.87 | .351 | .00 | [.00, .01] |
| Gender x Stereotype activation | 25.06 | 2 | 12.53 | 1.01 | .364 | .00 | [.00, .01] |
| Field of study/work x Stereotype activation | 21.41 | 2 | 10.71 | 0.86 | .422 | .00 | [.00, .01] |
| Gender x Field of study/work x Stereotype activation | 78.42 | 2 | 39.21 | 3.16 | .043 | .01 | [.00, .03] |
| Error | 8153.59 | 658 | 12.39 |  |  |  |  |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts.

Table S24.

*Fixed-Effects ANOVA results using political knowledge as the dependent variable (pooled sample)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 104552.99 | 1 | 104552.99 | 7777.11 | < .001 |  |  |
| Gender1 | 1710.26 | 1 | 1710.26 | 127.22 | < .001 | .08 | [.05, .11] |
| Field of study/work1 | 860.46 | 1 | 860.46 | 64.00 | < .001 |  .04 | [.02, .06] |
| Stereotype activation1 | 34.14 | 2 | 17.07 | 1.27 | .281 | .00 | [.00, .01] |
| Gender x Field of study/work | 15.50 | 1 | 15.50 | 1.15 | .283 | .00 | [.00, .01] |
| Gender x Stereotype activation | 61.85 | 2 | 30.93 | 2.30 | .101 | .00 | [.00, .01] |
| Field of study/work x Stereotype activation | 33.98 | 2 | 16.99 | 1.26 | .283 | .00 | [.00, .01] |
| Gender x Field of study/work x Stereotype activation | 8.51 | 2 | 4.25 | 0.32 | .729 | .00 | [.00, .00] |
| Error | 20031.08 | 1490 | 13.44 |  |  |  |  |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Factors were coded with orthogonal contrasts.

Table S25.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable and political interest as a covariate (pooled sample)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 102939.03 | 1 | 102939.03 | 8652.86 | < .001 |  |  |
| Political Interest1 | 2317.15 | 1 | 2317.15 | 194.78 | < .001 | .12 | [.09, .15] |
| Gender2 | 1248.41 | 1 | 1248.41 | 104.94 | < .001 | .07 | [.04, .09] |
| Field of study/work2 | 281.88 | 1 | 281.88 | 23.69 | < .001 | .02 | [.01, .03] |
| Stereotype activation2 | 27.92 | 2 | 13.96 | 1.17 | .310 | .00 | [.00, .01] |
| Gender x Field of study/work | 11.07 | 1 | 11.07 | 0.93 | .335 | .00 | [.00, .01] |
| Gender x Stereotype activation | 59.81 | 2 | 29.91 | 2.51 | .081 | .00 | [.00, .01] |
| Field of study/work x Stereotype activation | 30.85 | 2 | 15.43 | 1.30 | .274 | .00 | [.00, .01] |
| Gender x Field of study/work x Stereotype activation | 5.15 | 2 | 2.58 | 0.22 | .806 | .00 | [.00, .00] |
| Error | 17713.93 | 1489 | 11.90 |  |  |  |  |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts.

Table S26.

*Fixed-Effects ANCOVA results using frequency of “don’t know” responses as the dependent variable and political interest as a covariate*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 32422.65 | 1 | 32422.65 | 2279.09 | < .001 |  |  |
| Political interest1 | 5979.81 | 1 | 5979.81 | 420.34 | < .001 | .22 | [.18, .25] |
| Gender2 | 946.04 | 1 | 946.04 | 66.50 | < .001 | .04 | [.02, .06] |
| Stereotype activation2 | 36.18 | 2 | 18.09 | 1.27 | .281 | .00 | [.00, .01] |
| Gender x Stereotype activation | 75.51 | 2 | 37.76 | 2.65 | .071 | .00 | [.00, .01] |
| Error | 21268.11 | 1495 | 14.23 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts. |

Table S27.

*Fixed-Effects ANCOVA results using frequency of incorrect responses as the dependent variable and political interest as a covariate*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 48947.13 | 1 | 48947.13 | 9879.76 | < .001 |  |  |
| Political interest1 | 675.66 | 1 | 675.66 | 136.38 | < .001 | .08 | [.06, .11] |
| Gender2 | 226.00 | 1 | 226.00 | 45.62 | < .001 | .03 | [.02, .05] |
| Stereotype activation2 | 3.20 | 2 | 1.60 | 0.32 | .724 | .00 | [.00, .00] |
| Gender x Stereotype activation | 10.92 | 2 | 5.46 | 1.10 | .332 | .00 | [.00, .01] |
| Error | 7406.65 | 1495 | 4.95 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts. |

Table S28.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable, proportion of “don’t knows” to incorrect answers and political interest as a covariate*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 102298.84 | 1 | 102298.84 | 11967.66 | < .001 |  |  |
| Political interest1 | 231.31 | 1 | 231.31 | 27.06 | < .001 | .02 | [.01, .03] |
| Proportion “don’t know”1  | 4832.62 | 1 | 4832.62 | 565.35 | < .001 | .28 | [.24, .31] |
| Gender2 | 408.31 | 1 | 408.31 | 47.77 | < .001 | .03 | [.02, .05] |
| Field of study/work2 | 50.05 | 1 | 50.05 | 5.86 | .016 | .00 | [.00, .01] |
| Stereotype activation2 | 4.07 | 2 | 2.04 | 0.24 | .788 | .00 | [.00, .00] |
| Gender x Field of study/work | 2.50 | 1 | 2.50 | 0.29 | .589 | .00 | [.00, .00] |
| Gender x Stereotype activation | 8.62 | 2 | 4.31 | 0.50 | .604 | .00 | [.00, .00] |
| Field of study/work x Stereotype activation | 11.14 | 2 | 5.57 | 0.65 | .521 | .00 | [.00, .01] |
| Gender x Field of study/work x Stereotype activation | 0.90 | 2 | 0.45 | 0.05 | .949 | .00 | [.00, .00] |
| Error | 12702.24 | 1486 | 8.55 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts. |

Table S29.

*Fixed-Effects ANCOVA results using frequency of attempted responses in the political knowledge test as the dependent variable and political interest as a covariate*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 293751.53 | 1 | 293751.53 | 19387.92 | < .001 |  |  |
| Political interest1 | 5978.31 | 1 | 5978.31 | 394.57 | < .001 | .21 | [.17, .24] |
| Gender2 | 888.81 | 1 | 888.81 | 58.66 | < .001 | .04 | [.02, .06] |
| Stereotype activation2 | 38.55 | 2 | 19.27 | 1.27 | .281 | .00 | [.00, .01] |
| Gender x Stereotype activation | 59.18 | 2 | 29.59 | 1.95 | .142 | .00 | [.00, .01] |
| Error | 22651.14 | 1495 | 15.15 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts. |

Table S30.

*Fixed-Effects ANOVA results using political interest as the dependent variable*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 31204.88 | 1 | 31204.88 | 12300.85 | <.001 |  |  |
| Gender1 | 54.30 | 1 | 54.30 | 21.40 | <.001 | .01 | [.00, .03] |
| Field of study/work1 | 236.00 | 1 | 236.00 | 93.03 | <.001 | .06 | [.04, .08] |
| Stereotype activation1 | 5.25 | 2 | 2.62 | 1.04 | .355 | .00 | [.00, .01] |
| Gender x Field  | 0.61 | 1 | 0.61 | 0.24 | .625 | .00 | [.00, .00] |
| Gender x Stereotype activation | 1.86 | 2 | 0.93 | 0.37 | .693 | .00 | [.00, .00] |
| Field x Stereotype activation | 0.12 | 2 | 0.06 | 0.02 | .976 | .00 | [.00, 1.00] |
| Gender x Field x Stereotype activation | 1.28 | 2 | 0.64 | 0.25 | .777 | .00 | [.00, .00] |
| Error | 3779.84 | 1490 | 2.54 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Factors were coded with orthogonal contrasts. |

Table S31.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable and political interest as a covariate (Politics groups recoded)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 97023.65 | 1 | 97023.65 | 8184.77 | < .001 |  |  |
| Political interest1 | 2181.90 | 1 | 2181.90 | 184.06 | < .001 | .11 | [.08, .14] |
| Gender2 | 1091.11 | 1 | 1091.11 | 92.04 | < .001 | .06 | [.04, .08] |
| Field of study/work2 | 379.61 | 1 | 379.61 | 32.02 | < .001 | .02 | [.01, .04] |
| Stereotype activation2 | 21.19 | 2 | 10.60 | 0.89 | .409 | .00 | [.00, .01] |
| Gender x Field of study/work | 0.04 | 1 | 0.04 | 0.00 | .955 | .00 | [.00, 1.00] |
| Gender x Stereotype activation | 52.17 | 2 | 26.09 | 2.20 | .111 | .00 | [.00, .01] |
| Field of study/work x Stereotype activation | 12.79 | 2 | 6.39 | 0.54 | .583 | .00 | [.00, .00] |
| Gender x Field of study/work x Stereotype activation | 0.92 | 2 | 0.46 | 0.04 | .962 | .00 | [.00, .00] |
| Error | 17650.85 | 1489 | 11.85 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts. |

Table S32.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable and political interest as a covariate (Field of study/work not included)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 110523.78 | 1 | 110523.78 | 9149.85 | <.001 |  |  |
| Political interest1 | 2932.62 | 1 | 2932.62 | 242.78 | <.001 | .14 | [.11, .17] |
| Gender2 | 1469.80 | 1 | 1469.80 | 121.68 | <.001 | .08 | [.05, .10] |
| Stereotype activation2 | 19.28 | 2 | 9.64 | 0.80 | .450 | .00 | [.00, .01] |
| Gender x Stereotype activation | 48.21 | 2 | 24.11 | 2.00 | .136 | .00 | [.00, .01] |
| Error | 18058.55 | 1495 | 12.08 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts. |

Table S33.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable, political interest and ranking of Politics as covariates*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 101535.23 | 1 | 101535.23 | 8542.26 | <.001 |  |  |
| Political interest1 | 2133.38 | 1 | 2133.38 | 179.48 | <.001 | .11 | [.08, .14] |
| Ranking Politics1 | 27.22 | 1 | 27.22 | 2.29 | .130 | .00 | [.00, .01] |
| Gender2 | 1198.85 | 1 | 1198.85 | 100.86 | <.001 | .06 | [.04, .09] |
| Field of study/work2 | 138.25 | 1 | 138.25 | 11.63 | .001 | .01 | [.00, .02] |
| Stereotype activation2 | 27.45 | 2 | 13.72 | 1.15 | .315 | .00 | [.00, .01] |
| Gender x Field  | 7.08 | 1 | 7.08 | 0.60 | .440 | .00 | [.00, .00] |
| Gender x Stereotype activation | 61.64 | 2 | 30.82 | 2.59 | .075 | .00 | [.00, .01] |
| Field x Stereotype activation | 30.36 | 2 | 15.18 | 1.28 | .279 | .00 | [.00, .01] |
| Gender x Field x Stereotype activation | 5.02 | 2 | 2.51 | 0.21 | .810 | .00 | [.00, .00] |
| Error | 17686.71 | 1488 | 11.89 |  |  |  |  |
| *Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts. |

Table S34.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable, political interest and importance attributed to Politics as covariates*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 103152.35 | 1 | 103152.35 | 8874.20 | <.001 |  |  |
| Political interest1 | 2666.13 | 1 | 2666.13 | 229.37 | <.001 | .13 | [.10, .17] |
| Importance Of Politics1 | 417.65 | 1 | 417.65 | 35.93 | <.001 | .02 | [.01, .04] |
| Gender2 | 1248.88 | 1 | 1248.88 | 107.44 | <.001 | .07 | [.04, .09] |
| Field of study/work2 | 382.42 | 1 | 382.42 | 32.90 | <.001 | .02 | [.01, .04] |
| Stereotype activation2 | 30.88 | 2 | 15.44 | 1.33 | .265 | .00 | [.00, .01] |
| Gender x Field | 21.79 | 1 | 21.79 | 1.87 | .171 | .00 | [.00, .01] |
| Gender x Stereotype activation | 59.64 | 2 | 29.82 | 2.57 | .077 | .00 | [.00, .01] |
| Field x Stereotype activation | 36.84 | 2 | 18.42 | 1.58 | .205 | .00 | [.00, .01] |
| Gender x Field x Stereotype activation | 5.71 | 2 | 2.85 | 0.25 | .782 | .00 | [.00, .00] |
| Error | 17296.28 | 1488 | 11.62 |  |  |  |  |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts.

Table S35.

*Fixed-Effects ANCOVA results using political knowledge as the dependent variable, political interest and importance attributed to Politics as covariates on a young and educated subsample*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | SumofSquares | *df* | MeanSquare | *F* | *p* | partial η2 | partial η295% CI[LL, UL] |
| (Intercept) | 29329.24 | 1 | 29329.24 | 2712.76 | <.001 |  |  |
| Political interest1 | 394.31 | 1 | 394.31 | 36.47 | <.001 | .07 | [.03, .11] |
| Gender2 | 116.65 | 1 | 116.65 | 10.79 | .001 | .02 | [.00, .05] |
| Field of study/work2 | 89.23 | 1 | 89.23 | 8.25 | .004 | .02 | [.00, .04] |
| Stereotype activation2 | 15.56 | 2 | 7.78 | 0.72 | .487 | .00 | [.00, .02] |
| Gender x Field | 0.71 | 1 | 0.71 | 0.07 | .797 | .00 | [.00, .01] |
| Gender x Stereotype activation | 0.09 | 2 | 0.04 | 0.00 | .996 | .00 | [.00, 1.00] |
| Field x Stereotype activation | 13.82 | 2 | 6.91 | 0.64 | .528 | .00 | [.00, .02] |
| Gender x Field x Stereotype activation | 3.25 | 2 | 1.62 | 0.15 | .860 | .00 | [.00, .01] |
| Error | 5503.10 | 509 | 10.81 |  |  |  |  |

*Note.* LL and UL represent the lower-limit and upper-limit of the partial η2 confidence interval, respectively. 1 Covariates were mean-centered 2 Factors were coded with orthogonal contrasts.

**5. Reporting Standards**

**5.1 Hypotheses**

The present study aimed to replicate the results of study 2 of Ihme and Tausendpfund (2018), where it was found that activation of gender stereotypes affected individuals’ performance in a political knowledge test. The replication was preregistered and detailed hypotheses can be found at our preregistration (<https://osf.io/nxrg7>). As per the original study, we expected:

H1: a significant interaction between stereotype activation and gender for the political knowledge scores;

H2: a significant interaction between stereotype activation and gender for the political knowledge scores even after controlling for political interest;

H3a: male participants to have significantly higher scores in the political knowledge test than female participants in the stereotype activation conditions;

H3b: female participants to have significantly higher scores in the political knowledge test in the control condition compared to the stereotype activation by gender difference statement condition;

H3c: male participants to have significantly lower scores in the political knowledge test in the control condition compared to the stereotype activation by gender difference statement condition.

**5.2 Subjects and Context**

* *Eligibility and exclusion criteria for participants:* Only American citizens older than 18 years studying or working at the time of the survey were invited to take part in the study. As per our preregistration, we excluded participants who indicated they had cheated or used help to answer the political knowledge test; who failed attention check questions and who did not identify as male or female. Participants were recruited using a professional survey firm to ensure the sample size target could be reached. The criteria used for the composition of the Politics group slightly changed during data collection. We explain this change in detail in section 1.1 of the supplementary materials. This change was also amended to our preregistration (see Appendix A).
* *Procedures used to recruit and select participants:*Participants were recruited using a professional survey firm (<https://www.cint.com>). Cint has the world’s largest network of integrated survey panels (4,500+) and applies a variety of industry-standard 3rd party solutions to ensure data quality.
* *Recruitment dates:*Stage 1 (29th October 2020 to 4th November 2020); Stage 2 (5th November 2020 to 10th November 2020).
* *Settings and locations where the data were collected:*Data were collected onlineusing the Qualtrics survey platform.
* *Response rate:*Because we collected data via a professional survey firm, we do not have the necessary information to calculate the response rate.

**5.3 Allocation Method**

* *Details of the procedure used to generate the assignment sequence:* Participants were

randomly assigned (via Qualtrics randomization) to one of the three stereotype activation conditions. In addition, participants were assigned to either the Politics or non-Politics group based on their responses (see section 1.1 of the Supplementary Materials).

* *Details of randomization procedure:* Randomization of participants to the three stereotype activation conditions was done at the individual level.
* *Evidence of random assignment:* Tables S2 and S5 in the Supplementary Materials show the distribution of participants according to gender and field of work/study (Politics/Non-Politics) for each of the stereotype activation conditions.
* *Blinding:* Participants were not aware of the treatment group they were assigned to. In addition, because the study was conducted online, researchers did not directly interact with participants.
	1. **Treatments**
* *Description of the interventions in each treatment condition, as well as a description of the control group:* A detailed description of the stereotype activation conditions is provided in section 1.3 of the Supplementary Materials. Instructions for the political knowledge test in each stereotype activation condition are provided below:

Stereotype not activated (control condition):

Below you will be given a number of questions about your political knowledge. Please answer these questions as well as possible. Do not use any assistance (e.g., checking the internet or asking others). Every question will be displayed for up to 20 seconds. However, you can use the "continue" button to move ahead when you are done. If you do not know the answer to a certain question, mark "don't know" and click "continue".

Stereotype activated by gender question:

Same instructions as for the control condition above. However, immediately before receiving the instructions for the political knowledge test, participants were asked to report their gender:

Please indicate your gender: Female/Male/Other (intersexual, transsexual etc).

Stereotype activated by gender question:

Below you will be given a number of questions about your political knowledge, that have been known to produce gender differences with regard to the number of correct answers. Please answer these questions as well as possible. Do not use any assistance (e.g., checking the internet or asking others). Every question will be displayed for up to 20 seconds. However, you can use the "continue" button to move ahead when you are done. If you do not know the answer to a certain question, mark "don't know" and click "continue".

* *How and when manipulations or interventions were administered:*

-Method of delivery: online (computer, smartphone, tablets)

-Software: Qualtrics (the .qsf file can be found at the OSF project folder)

-Additional information: Participants could only participate in the experiment once. There was no repetition of experimental conditions and no within-subject factors. The political knowledge test was only completed once per participant. As in the original study, there were no practice rounds of the political knowledge test nor manipulation checks of the stereotype activation conditions. No deception was employed and participants were debriefed on the study goals when they completed the study. Participants took on average 8.95 minutes (*SD* = 9.65) to complete the study.

**5.5 Results**

**5.5.1 Outcome Measures and Covariates**

* A detailed description of all measures in the study can be found in section 1 of the Supplementary Materials. The complete questionnaire can be found in Appendix C.
* Exploratory analyses are labelled as such in the manuscript.

**5.5.2 CONSORT Participant Flow Diagram**



**Figure S2.** CONSORT participant flow diagram for Stage 1.



**Figure S3.** CONSORT participant flow diagram for Stage 2.

**5.5.3 Statistical Analysis**

The distribution of political knowledge scores, as well as unconditional means per experimental condition can be seen in Figure 1 in the manuscript. In addition, we report means and standard deviations of “don’t know”, incorrect and attempted responses in the political knowledge test per experimental condition in tables S9, S11 and S14, respectively.

**5.6 Other Information**

* This study was approved by an independent IRB ethics committee, BRANY (https://www.brany.com) and the U.S. Army’s Human Research Protection Office (HRPO).
* The experimental design, hypotheses and analyses plan were preregistered at OSF (<https://osf.io/nxrg7>)
* This work was carried out as part of the Center for Open Science’s Systematizing Confidence in Open Research and Evidence (SCORE) program, which is funded by the Defense Advanced Research Projects Agency. The funders had no role in planning the study or analyzing the results.
* Replication materials can be found at: <https://doi.org/10.7910/DVN/ETUUOD>

**References for the Supplementary Materials**

Annenberg Public Policy Center. 2019. *2019 Annenberg Civics Knowledge Constitution Day Survey*. Retrieved from <https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2019/09/Annenberg_civics_2019_Appendix.pdf>

Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. 2018. Emmeans: Estimated marginal means, aka least-squares means. *R package version, 1*(1), 3.

Mondak, J. J. 1999. Reconsidering the measurement of political knowledge. *Political Analysis, 8*(1), 57-82. <https://doi.org/10.1093/oxfordjournals.pan.a029805>

Mondak, J. J., and Anderson, M. R. 2004. The knowledge gap: A reexamination of gender-based differences in political knowledge. *The Journal of Politics*, *66*(2), 492-512. <https://doi.org/10.1111/j.1468-2508.2004.00161.x>

Otto, L. and Bacherle, P. 2011. Politisches Interesse Kurzskala (PIKS) – Entwicklung und Validierung. *Politische Psychologie, 1*(1), 19–35.

1. Identical to the original study, we also conducted one-tailed *t*-tests. No difference in political knowledge test performance could be found between women in the stereotype not activated and the stereotype activated by gender-question, *t*(486) = -1, *p* = .08, *d* = -.13, or gender difference statement condition, *t*(465) = -1, *p* = .09, *d* = -.13. A similar pattern was found for men; their performance in the stereotype not activated condition was not significantly different from their performance in the stereotype activated by gender question, *t*(516) = -.30, *p* = .6, *d* = -.02, or gender difference statement condition, *t*(512) = 1, *p* = .1, *d* = .11. [↑](#footnote-ref-1)