Replication of a Research Claim from Ihme & Tausendpfund (2018)
from The Journal of Experimental Political Science

Replication Team: Flávio Azevedo, Deliah Bolesta, Leticia Micheli

Research Scientist: Andrew Tyner

Action Editor: Kimberly Quinn

Independent Reviewers
(add name below when you initiate review, comment “DONE” on your name when you finish):

Reviewer #1: Landon Schnabel

Reviewer #2: [NAME]

Reviewer #3: [NAME]

Review Period: Feb. 3 - Feb. 10

View-only links to: Original Paper, Original Materials

Note to Reviewers: Going forward, we will be implementing a new preregistration form. Please reference the New Preregistration Form and the Reviewer Criteria Checklist for instructions and specific criteria to help guide your review.

Privacy Statement: Other teams are making predictions about the outcomes of many different studies, not knowing which studies have been selected for replication. As a consequence, the success of this project requires full confidentiality of this peer review process. This includes privacy about which studies have been selected for replication and all aspects of the discussion about these replication designs.

General information about preregistration is available at https://cos.io/prereg and you can reach out to the SCORE coordinators at scorecoordinator@cos.io or make a comment in the document for additional assistance.

Sourcing Notes

- The original study was conducted in Germany. The corresponding authors had the following to say about a replication attempt that would take place outside of Germany.
    - "The key to a replication of our study – no matter whether with German or non-German participants – will be the political knowledge test. However, to our knowledge, there is no standardized test for political knowledge. Considering the fluid nature of politics and therefore political knowledge (today's ministers might be unimportant and without a position tomorrow; laws and political structures change and so on) every study investigating stereotype threat and political knowledge will need a specific test relevant for the point in time and the sample (for example German politics, British politics and so on). Furthermore, one of the basic conditions of a stereotype threat effect is a sufficiently difficult test. That means, for a successful replication you need to create difficult test items."

Study Information

---

1. Title
    - *1.1. Provide the working title of your study. It may be the same title that you submit for publication of your final manuscript, but it is not a requirement.*
    - *1.2.* ***Example****: Effect of sugar on brownie tastiness.*
    - *1.3.* ***More info****: The title should be a specific and informative description of a project. Vague titles such as 'Fruit fly preregistration plan' are not appropriate.*

TITLE: Direct replication of a research claim from Ihme & Tausendpfund (2018), in Journal of Experimental Political Science

---

2. Authors (required)

AUTHORS: Flavio Azevedo[1], Deliah Bolesta, Leticia Micheli

RR LAB LEAD[1]

3. Description
    3.1. Please give a brief description of your study, including some background, the purpose of the of the study, or broad research questions.
    3.2. ***Example****: Though there is strong evidence to suggest that sugar affects taste preferences, the effect has never been demonstrated in brownies. Therefore, we will measure taste preference for four different levels of sugar concentration in a standard brownie recipe to determine if the effect exists in this pastry.*
    3.3. ***More info****: The description should be no longer than the length of an abstract. It can give some context for the proposed study, but great detail is not needed here for your preregistration.*

DESCRIPTION:

The claim selected for replication from Ihme & Tausendpfund (2018) is that the activation of the gender stereotype affects female participants' performance on a political knowledge test, which reflects the following statement from the abstract: "...the activation of gender stereotypes affects performance on a political knowledge test." The study conducted a 2 × 2 × 3 ANCOVA for the dependent variable (political knowledge test score) using political interest as the covariate. The independent variables were the participants' gender (male vs. female), the participants' field of study (psychology vs. politics), and stereotype activation (stereotype not activated vs. stereotype activated by gender-question vs. stereotype activated by gender difference statement). Subjects were randomly assigned to the respective gender stereotype conditions. Controlling for political interest, the interaction of stereotype activation and gender ($F(2,364) = 6.17$, $p = 0.002$, $\eta^2 = 0.03$) was significant.

4. Hypotheses
    - *4.1. List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.*
    - *4.2.* ***Example****: If taste affects preference, then mean preference indices will be higher with higher concentrations of sugar.*

HYPOTHESIS:

H1a: As in the original study, we expect to find a significant interaction between stereotype activation (stereotype not activated vs. stereotype activated by gender question vs. stereotype activated by gender difference statement) and gender (male vs female) for the political knowledge test score outcomes.

H2: We expect to still find a significant interaction between stereotype activation and gender for the political knowledge test score outcomes when controlling for political interest.

H3a: Male participants are expected to have significantly higher scores in the political knowledge test than female participants in the stereotype activated by gender question condition and in the stereotype activated by gender difference statement condition. No differences between female and male participants are expected in the stereotype not activated condition.

H3b: Female participants are expected to have significantly higher scores in the political knowledge test in the stereotype not activated condition than in the stereotype activated by gender difference statement condition. No differences in female participants' scores are expected between the stereotype not activated and the stereotype activated by gender question conditions.

H3c: Male participants are expected to have significantly lower scores in the political knowledge test in the stereotype not activated condition than in the stereotype activated by gender difference statement condition. No differences in male participants' scores are expected between the stereotype not activated and the stereotype activated by gender question conditions.

Design Plan

In this section, you will be asked to describe the overall design of your study. Remember that this research plan is designed to register a single study, so if you have multiple experimental designs, please complete a separate preregistration.

---

5. *Study type*
    - *5.1. Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.*
    - *5.2. Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.*
    - *5.3. Meta-Analysis - A systematic review of published studies.*
    - *5.4. Other*

STUDY TYPE: Experimental study

6. Blinding (required)
    - *6.1. Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.*
        - *6.1.1. No blinding is involved in this study.*
        - *6.1.2. For studies that involve human subjects, they will not know the treatment group to which they have been assigned.*
        - *6.1.3. Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)*
        - *6.1.4. Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.*

7. Is there any additional blinding in this study?

BLINDING:
Participants will not be aware of the treatment group to which they have been assigned. As the study will be conducted online, there will be no personnel directly interacting with participants during the experiment.

---

8. Study design
    8.1. Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.
    8.2. Example: We have a between subjects design with 1 factor (sugar by mass) with 4 levels.
    8.3. More info: This question has a variety of possible answers. The key is for a researcher to be as detailed as is necessary given the specifics of their design. Be careful to determine if every parameter has been specified in the description of the study design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. For example, if the study design describes a complete factorial, 2 X 3 design and the treatments and levels are specified previously, you do not have to repeat that information.

STUDY DESIGN: This study uses a 2 (gender [male vs. female]) × 2 (field of study/work [non-politics vs. politics]) × 3 (stereotype activation [stereotype not activated vs. stereotype activated by gender-question vs. stereotype activated by gender difference statement]) between subjects design, where stereotype activation is manipulated via survey instructions.

Please note: As the claim we are attempting to replicate only concerns the interaction between gender and stereotype activation, we do not include the variable “field of study/work” in any of the hypothesis described above. However, we still include this variable in the study design, as we would like to run a 2x2x3 ANCOVA for the dependent variable (political knowledge test score), just like in the original study. This is done as an attempt to have the replication design and analyses as close as possible to the original study.

---

9. Randomization (required)
    9.1. If you are doing a randomized study, how will you randomize, and at what level?
    9.2. Example: We will use block randomization, where each participant will be randomly assigned to one of the four equally sized, predetermined blocks. The random number list used to create these four blocks will be created using the web applications available at http://random.org.
    9.3. More info: Typical randomization techniques include: simple, block, stratified, and adaptive covariate randomization. If randomization is required for the study, the method should be specified here, not simply the source of random numbers.

RANDOMIZATION:

Participants will be randomly allocated into one of the three stereotype activation conditions. Randomization will be done via Qualtrics, which has a randomization feature that ensures that each participant is randomly assigned to one of the stereotype activation conditions.

Though not the focal randomization for the experiment, the items of the political knowledge test will be presented in random order, as in the original study. This randomization will also be done in Qualtrics via Qualtrics features. _____

Sampling Plan

In this section we'll ask you to describe how you plan to collect samples, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.

---

10. Existing data
    - *10.1. Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges.* ***Please select the description that best describes your situation.*** *Please do not hesitate to contact us if you have questions about how to answer this question (prereg@cos.io).*
        - 10.1.1. Registration prior to creation of data: As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized.

---

11. Explanation of existing data
    - *11.1. If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.*
    - *11.2.* ***Example****: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.*
    - *11.3.* ***More info****: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.*

EXPLANATION OF EXISTING DATA: N/A

12. Data collection procedures

12.1. Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that donÍt include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.

*12.2. **Example**: Participants will be recruited through advertisements at local pastry shops. Participants will be paid $10 for agreeing to participate (raised to $30 if our sample size is not reached within 15 days of beginning recruitment). Participants must be at least 18 years old and be able to eat the ingredients of the pastries.*

*12.3. **More information**: The answer to this question requires a specific set of instructions so that another person could repeat the data collection procedures and recreate the study population. Alternatively, if the study population would be unable to be reproduced because it relies on a specific set of circumstances unlikely to be recreated (e.g., a community of people from a specific time and location), the criteria and methods for creating the group and the rationale for this unique set of subjects should be clear.*

DATA COLLECTION PROCEDURES:

In this replication attempt, we try to keep as much as possible the same experimental design of the original study. Necessary changes to the original design are mentioned when appropriate.

667 participants will be recruited via a professional survey company (i.e., CINT) and will be paid a fee for completing the study. The precise fee will be established by the professional survey company for the duration of the study. The replication will be conducted with American participants, given the online panels the researchers have at their disposal contain a higher proportion of Americans compared to other nationalities (e.g., Germans - as in the original study). The professional survey companies will be instructed to only invite participants who are American citizens, older than 18 years old and who are studying or working. Participants who do not meet this criteria will not be invited to take part in the study.

The call for quote proposals is ongoing among survey companies. However we have received a satisfactory answer from CINT, which can only deliver the ‘Politics’ group (see below) on a best effort basis. Hence, iIt is likely that more than one survey company will be necessary to achieve the desired sample size for the ‘Politics’ group. For this reason we have secured (and will continue to pursue) quote proposals for additional professional survey companies (Lucid, Dynata). We will activate them only if necessary.

Upon providing consent, participants will be asked to indicate a maximum of 4 topics (from a comprehensive list containing 28 topics) that are relevant for their studies or for their work. The list will contain the following topics: *Politics, Psychology, Economics, Biology, History, Geography, Chemistry, Physics, Business, Food industry, Sociology, Public administration, Mathematics, Law, Accounting, Engineering, Pedagogy, Education, Health Care, Social Services, Arts, Finance, Transportation, Retail, Information Technology, Real estate, Government, Other.* This will be our manipulation of field of study. Participants who indicate Politics is important for their study/work will compose the Political group, whereas participants who do not indicate Politics among their four selected topics will compose the Non-Political group. We aim to have groups with the same proportion of participants as in the original study.We will control the ratio of participants in each group by means of quotas implemented on Qualtrics. After the quota of a specific group has been met, participants who do not fit the selection criteria will be redirected to the end of the survey and will be paid a reduced fee. We note that, differently than the original study, in this replication we ask for topics relevant to their field of study/work, as it would be challenging to find the target number of Politics/Psychology students in online panels. For the same reason, we broaden the manipulation to field of study and/or work, as to be able to also include non-students in the sample, given that online panels are mostly composed of working young adults.

Participants will then be asked to fill in the short scale of Political Interest. We will use the same scale as in the original paper. However, we will use the items in English as validated by Otto and Bacherle (2011). We will then instruct participants to answer a political knowledge test, adapted to the context and reality of current American politics. Participants will be randomly allocated into the three stereotype manipulation conditions, which will be the same of the original study. That is, depending on the experimental condition they read either gender-wise neutral instructions for the political knowledge test (stereotype not activated), gender-wise neutral instructions for the political knowledge test but are prompted to indicate their gender (stereotype activated by a gender-question), or instructions indicating gender differences in the political knowledge test (stereotype activated by a gender difference statement).

Following the instructions, participants will answer the political knowledge test, which will be developed to be relevant for the American population. Items will be presented in a random order and will have a time limit and a “don’t know” option, as in the original study. We will then include the same control questions as in the original study (i.e., asking if participants answered the questions without help and without cheating) and demographic questions (i.e., age, education, field of study/work, and gender - for the conditions which did not require stereotype activation by the gender question). We will also include two additional control questions in order to have a more fine grained understanding of how much Politics is important for participants’ study/work. We will ask participants to rank the topics presented earlier in order of importance to their study/work as well as ask participants the extent to which each topic from the list presented earlier in the experiment is important for their study/work(all responses should sum up to 100).

13. Sample size

13.1. *Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?*

13.2. ***Example**: Our target sample size is 280 participants. We will attempt to recruit up to 320, assuming that not all will complete the total task.*

13.3. ***More information**: For some studies, this will simply be the number of samples or the number of clusters. For others, this could be an expected range, minimum, or maximum number.*

13.4. *This could include a power analysis or an arbitrary constraint such as time, money, or personnel.*

13.5. ***Example**: We used the software program G*Power to conduct a power analysis. Our goal was to obtain .95 power to detect a medium effect size of .25 at the standard .05 alpha error probability.*

13.6. ***More information**: This gives you an opportunity to specifically state how the sample size will be determined. A wide range of possible answers is acceptable; remember that transparency is more important than principled justifications. If you state any reason for a sample size upfront, it is better than stating no reason and leaving the reader to "fill in the blanks." Acceptable rationales include: a power analysis, an arbitrary number of subjects, or a number based on time or monetary constraints.*

SAMPLE SIZE:

The initial target sample size is 667 participants. If a statistically significant effect is not observed after the first round of data collection, a second round will begin. The second round of data collection will sample an additional 830 participants for a pooled sample of 1497 participants**.**

SAMPLE SIZE RATIONALE:

Sample size rationale

Power calculations were done in accordance with the guidelines of the Social Sciences Replication Project (SSRP). The first round of data collection achieves 90% power to detect 75% of the original effect size. The pooled sample, if necessary after testing the effect on the first round of data, achieves 90% power to detect 50% of the original effect size.

The power analysis was performed by the Center for Open Science SCORE team using the pwr package in R, using a script for an ANCOVA. The R markdown file that produced the analysis is contained here.

14. Stopping rule
    - *14.1. If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.*
    - *14.2.* ***Example****: We will post participant sign-up slots by week on the preceding Friday night, with 20 spots posted per week. We will post 20 new slots each week if, on that Friday night, we are below 320 participants.*
    - *14.3.* ***More information****: You may specify a stopping rule based on p-values only in the specific case of sequential analyses with pre-specified checkpoints, alphas levels, and stopping rules. Unacceptable rationales include stopping based on p-values if checkpoints and stopping rules are not specified. If you have control over your sample size, then including a stopping rule is not necessary, though it must be clear in this question or a previous question how an exact sample size is attained.*

STOPPING RULE:

The planned sample size is 667 participants, that is, complete observations (i.e., participants have successfully gone through the survey and passed data-quality checks at both the survey company level and our own data quality checks). After achieving that sample, planned analyses will be run. If a significant effect is found, sampling stops. If a significant effect is not found, a second round of data collection will collect data from 830 additional participants, for a pooled sample of 1497 participants. Sampling will stop after the second round of data collection regardless of a significant effect.

Variables

In this section you can describe all variables (both manipulated and measured variables) that will later be used in your confirmatory analysis plan. In your analysis plan, you will have the opportunity to describe how each variable will be used. If you have variables which you are measuring for exploratory analyses, you are not required to list them, though you are permitted to do so.

---

15. Manipulated variables
    - *15.1. Describe all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.*
    - *15.2.* ***Example:*** *We manipulated the percentage of sugar by mass added to brownies. The four levels of this categorical variable are: 15%, 20%, 25%, or 40% cane sugar by mass.*
    - *15.3.* ***More information****: For any experimental manipulation, you should give a precise definition of each manipulated variable. This must include a precise description of the levels at which each variable will be set, or a specific definition for each categorical treatment. For example, "loud or quiet," should instead give either a precise decibel level or a means of recreating each level. 'Presence/absence' or 'positive/negative' is an acceptable description if the variable is precisely described.*

MANIPULATED VARIABLES:

Stereotype activation

We manipulate stereotype activation in the same way that the authors of the original study did, using the test instructions to produce the stereotype activation manipulation available in the supplementary materials and on the OSF here.

In the control condition (stereotype not activated) there is no mention of gender before the test. In the stereotype activated by a gender question, the manipulation is achieved by asking participants to indicate their gender before the test. Finally, in the stereotype activated by a gender difference statement condition, the instruction of the test mentions that the political knowledge test participants are about to answer has shown gender differences in the past. As in the original study, the direction of these differences are not mentioned.

---

16. Measured variables

16.1. Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.

16.2. ***Example****: The single outcome variable will be the perceived tastiness of the single brownie each participant will eat. We will measure this by asking participants 'How much did you enjoy eating the brownie' (on a scale of 1-7, 1 being 'not at all', 7 being 'a great deal') and 'How good did the brownie taste' (on a scale of 1-7, 1 being 'very bad', 7 being 'very good').*

16.3. ***More information****: Observational studies and meta-analyses will include only measured variables. As with the previous questions, the answers here must be precise. For example, 'intelligence,' 'accuracy,' 'aggression,' and 'color' are too vague. Acceptable alternatives could be 'IQ as measured by Wechsler Adult Intelligence Scale' 'percent correct,' 'number of threat displays,' and 'percent reflectance at 400 nm.'*

MEASURED VARIABLES:

Short Scale Political Interest

Just like in the original study, after reading an introduction and giving their consent the participants will be asked to answer the Short Scale Political Interest. Items will be presented in English. As in the original study, the scale will be answered on a 7-point scale ranging from 1 (not true at all) to 7 (completely true).

Political Knowledge

For the replication study we will use a political knowledge test about the USA, since our participants will be American. Just as in the original study, to ensure an adequate level of difficulty, we will include items about politicians who are less known or who get less coverage in the media. We will also give participants a time limit to answer each item, as in the original study.

Field of study/work

We ask participants to select from an extensive list a maximum of 4 topics which are relevant for their studies (in case participants are students) or for their work (in case participants are no longer students). Participants who select "Politics" will be in the Political group, while participants who do not select Politics among their 4 choices will be in the Non-political group.

Gender

We ask participants to report their gender either in the beginning or end of the study (depending on the stereotype condition they are allocated to). By means of quotas implemented on qualtrics, we aim to achieve a gender balanced sample.

Cheating

We will ask participants "Did you answer the questions without assistance?" The response categories will be Yes or No.

Rationale:

Whenever possible, the German items were translated in a way that is still meaningful and representative for a general political knowledge. All of those items (except for two) will be part of the Political Knowledge scale administered in this study. Items 4 and 5 were removed as the Green Party in Germany is much more popular among the general public than their equivalent in the US.

Item 9 and items 15 to 20 were removed because they tackled specific political affairs in Germany for which an exact equivalent in an US American context is absent or hard to find. Therefore, nine items from a well known and validated scale (the Annenberg Civics Knowledge Survey) will be included with slight modifications. Q1 was excluded because of the binary response option. Q2 was rephrased so that the response option could be multiple-choice instead of open-ended.

Any item **to be included** in the survey has the number marked in **bold** font.

Items of Political Knowledge Test in Ihme & Tausendpfund (2018) as listed in their appendix.

Dimension	#	Item wording	Notes	American equivalent	Multiple choice options (correct answer in *italics*)
Party allegiances of politicians and names of certain Federal ministers	**1**	Heiko Maas is member of the following party:	Heiko Maas: Minister of Foreign Affairs (party: SPD)	**Mike Pompeo is a member of which party:**	a) Democratic Party *b) Republican Party* c) Libertarian Party *d)* Green Party e) Don't know
	2	Andrea Nahles is member of the following party:	Andrea Nahles: former Leader of the SPD	**Nancy Pelosi is a member of which party:**	*a) Democratic Party* b) Republican Party c) Libertarian Party *d)* Green Party e) Don't know
	3	Gerd Müller is member of the following party:	Gerd Müller: Minister of Economic Cooperation and Development (party: CSU)	**Wilbur Ross is a member of which party:**	a) Democratic Party *b) Republican Party* c) Libertarian Party *d)* Green Party

					e) Don't know
	4	Anton Hofreiter is member of the following party:	Anton Hofreiter: Co-Leader of Bündnis 90/Die Grünen	Jill Stein is a member of which party:	a) Democratic Party b) Republican Party c) Libertarian Party *d) Green Party* e) Don't know
	5	Katrin Göring-Eckhardt is member of the following party:	Katrin Göring-Eckhardt: Co-Leader of Bündnis 90/Die Grünen	Ajamu Baraka is a member of which party:	a) Democratic Party b) Republican Party c) Libertarian Party *d) Green Party* e) Don't know
	6	Who is the current Federal minister for family, senior citizens, women and the young?	Franziska Giffey (party: SPD)	**Who is the current Secretary of Labor?**	a) Steven Mnuchin b) Elaine Chao c) Alex Azar *d) Eugene Scalia* e) Don't know
Details of the German political system	**7**	Who is the current Federal minister for health?	Jens Spahn (party: CDU)	**Who is the current Secretary of Health and Human Services?**	*a) Alex Azar* b) Mark Esper c) Eugene Scalia d) William Barr e) Don't know
	8	Who elects the Federal Chancellor of Germany?	Elected by a majority of the members of the Bundestag upon the proposal of the President	**Who elects the President of the United States?**	a) The Congress *b) The Electoral College* c) The National Convention d) The Political Parties e) Don't know
	9	Who commands the so-called 'Richtlinienkompetenz'?	The German Chancellor is authorized to define the course of politics to be followed irrespective of any potential disagreement with their Federal Ministers in this regard.	N/A	N/A

	10	Who elects the Federal President of Germany?	The Federal Convention	**Who elects the Vice President of the United States?**	a) The Congress *b) The Electoral College* c)The National Convention d) The Political Parties e) Don't know
	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?	The second vote where electors vote for a party in general, i.e. a regional electoral list put together by each party. The first vote determines who will get a "direct mandate", i.e. candidates can be voted for directly.	What is the so called "winner takes all" principle in an electoral system?	*a) the candidate who receives the relative majority of votes wins, even if they did not receive an absolute majority of the votes* b) the candidate who receives the absolute majority of votes wins, even if they did not receive a relative majority of the votes c) the candidate who receives the majority of votes gets to appoint the Senators. d) The candidate who receives the majority of votes gets to appoint the members of the House of Representatives. e) Don't know
Details on the current state of affairs in Germany	**12**	How high was the percentage of foreign nationals in Germany at the end of 2013?	7.015.000 (according to statista)	In 2019, what percentage of the American population was foreign-born?	a) 26.9% *b) 13.7%* c) 8.2% d) 3.7% e) Don't know
	13	What is the current (1. quarter 2015) unemployment rate in Germany?	6.4% in all of 2015 (according to bpb)	In 2019, what was the annual unemployment rate in the United States?	a) 26.9% b) 13.7% c) 8.2% *d) 3.7%* e) Don't know

	14	Which political department has the highest budget?	The Federal Ministry of Labor and Social Affairs	Which federal executive department of the United States has the highest budget at their disposal?	*a) The Secretary of Health and Human Services* b) The Secretary of Labor c) The Secretary of Defense d) The Secretary of Education e) Don't know
	15	Which one of the following parties argued for the collection and storage of contact data (e.g. telephone, internet) even without concrete cause?		N/A	N/A
	16	Which one of the following parties argued against raising the top income tax rate?		N/A	N/A
	17	Which one of the following parties argued for a general speed limit on the autobahn?		N/A	N/A
	18	Which one of the following parties argued against BAföG-payments regardless of parents' income?	BAföG = Germany's Federal Training Assistance Act for students at secondary schools and universities in Germany	N/A	N/A
	19	Which one of the following parties argued for a stricter public control of electricity tariffs?		N/A	N/A
	20	Which one of the following parties argued against a decrease of the statutory retirement age?		N/A	N/A

Translation of general terms and names (alphabetical order):

Bundestag – German Federal Parliament

Bündnis 90/Die Grünen – Green party

CDU – Christian Democratic Party

CSU – Bavarian sister party of the CDU

Die Linke – The Left Party

FDP – Free Democratic Party

SPD – Social Democratic Party

Annenberg Civics Knowledge Survey 2019

#	Item	Multiple choice options (correct answer in *italics*)
Q1	Do you happen to know any of the three branches of government?	a) Yes b) No c) Don't know
Q2	**Which of the following does not describe one of the three branches of government?** (Original wording: Would you mind naming any of them, the three branches of government?)	a) Judicial *b) Governmental* c) Executive d) Legislative e) Don't know
Q3	**Which political party is in control of the U.S. House of Representatives, or are you not sure?**	*f) Democratic Party* g) Republican Party h) Libertarian Party i) Green Party j) Don't know
Q4	**Which political party is in control of the U.S. Senate, or are you not sure?**	a) Democratic Party *b) Republican Party*

		c) Libertarian Party d) Green Party e) Don't know
Q5	**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?**	a) The President b) The U.S. Congress *c) The U.S. Supreme Court* d) The U.S. Senate e) Don't know
Q6	**How much of a majority is required for the U.S. Senate and the House of Representatives to override a presidential veto?**	a) 51 percent *b) Two-thirds* c) Three-quarters d) 90 percent e) Don't know
Q7	**If the U.S. Supreme Court rules on a case 5 to 4, does this mean …**	*a) The decision is the law of the land and needs to be followed* b) The decision is sent back to Congress for reconsideration c) The decision is sent back to the federal court of appeals to be decided there d) The U.S. Supreme Court needs to reconsider and come to a unanimous decision e) Don't know
Q8	**How accurate is the following statement: "the U.S. Supreme Court has held that a citizen has a constitutional right to own a handgun"?**	*a) Accurate* b) Neither accurate nor inaccurate c) Inaccurate d) Don't know
Q9	**How accurate is the following statement: "the U.S. Constitution allows a judge to insist that a defendant testify at his own trial"?**	a) Accurate b) Neither accurate nor inaccurate *c) Inaccurate*

		d) Don't know
Q10	**How accurate is the following statement: "those who are in the country illegally do not have any rights under the U.S. Constitution"?**	a) Accurate b) Neither accurate nor inaccurate *c) Inaccurate* d) Don't know

17. Indices

17.1. If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If your are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.

17.2. ***Example****: We will take the mean of the two questions above to create a single measure of 'brownie enjoyment.'*

17.3. ***More information****: If you are using multiple pieces of data to construct a single variable, how will this occur? Both the data that are included and the formula or weights for each measure must be specified. Standard summary statistics, such as "means" do not require a formula, though more complicated indices require either the exact formula or, if it is an established index in the field, the index must be unambiguously defined. For example, "biodiversity index" is too broad, whereas "Shannon's biodiversity index" is appropriate.*

INDICES:

As in the original study, for the Short Scale Political Interest, we will calculate a single index score per participant by averaging responses over the items. Additionally, we will provide results for each scale.

The dependent variable will be the sum of the political knowledge questions answered correctly.

Analysis Plan

You may describe one or more confirmatory analysis in this preregistration. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis generating.

A confirmatory analysis plan must state up front which variables are predictors (independent) and which are the outcomes (dependent), otherwise it is an exploratory analysis. You are allowed to describe any exploratory work here, but a clear confirmatory analysis is required.

---

18. Statistical models
    - *18.1. What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors, outcomes, or covariates). Please specify any interactions, subgroup analyses, pairwise or complex contrasts, or follow-up tests from omnibus tests. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Remember that any test not included here must be noted as an exploratory test in your final article.*
    - ***18.2. Example**: We will use a one-way between subjects ANOVA to analyze our results. The manipulated, categorical independent variable is 'sugar' whereas the dependent variable is our taste index.*
    - *18.3. **More information**: This is perhaps the most important and most complicated question within the preregistration. As with all of the other questions, the key is to provide a specific recipe for analyzing the collected data. Ask yourself: is enough detail provided to run the same analysis again with the information provided by the user? Be aware for instances where the statistical models appear specific, but actually leave openings for the precise test. See the following examples:*
        - *18.3.1.1. If someone specifies a 2x3 ANOVA with both factors within subjects, there is still flexibility with the various types of ANOVAs that could be run. Either a repeated measures ANOVA (RMANOVA) or a multivariate ANOVA (MANOVA) could be used for that design, which are two different tests.*
        - *18.3.1.2. If you are going to perform a sequential analysis and check after 50, 100, and 150 samples, you must also specify the p-values you'll test against at those three points.*

STATISTICAL MODELS:
For the claim we intend to replicate, we will run a 2 (Gender: male, female) x 2 (Field of work/study: Politics vs non-politics) x 3 (Stereotype activation: stereotype not activated,

stereotype activated by gender-question, stereotype activated by gender difference statement) ANCOVA controlling for political interest. The dependent variable is the individual score in the political knowledge test.

- The specific analysis selected is the interaction conveyed in the following: “When we controlled for political interest, the effects of gender ($F(1,364) = 19.23$, $p < 0.001$, $\eta^2 = 0.05$), field of study ($F(1,364) = 24.57$, $p < 0.001$, $\eta^2 = 0.06$), and the interaction of stereotype activation and gender ($F(2,364) = 6.17$, $p = 0.002$, $\eta^2 = 0.03$) remained significant.”

Data and syntax from the original authors is to be found here: https://doi.org/10.7910/DVN/OZRQIQ”

---

19. Transformations (optional)
    - *19.1. If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.*
    - *19.2.* ***Example****: The "Effect of sugar on brownie tastiness" does not require any additional transformations. However, if it were using a regression analysis and each level of sweet had been categorically described (e.g. not sweet, somewhat sweet, sweet, and very sweet), 'sweet' could be dummy coded with 'not sweet' as the reference category.*
    - *19.3.* ***More information****: If any categorical predictors are included in a regression, indicate how those variables will be coded (e.g. dummy coding, summation coding, etc.) and what the reference category will be.*

TRANSFORMATIONS:

_____

20. Inference criteria

20.1. What criteria will you use to make inferences? Please describe the information youÍll use (e.g. p-values, bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?

*20.2. **Example**: We will use the standard p<.05 criteria for determining if the ANOVA and the post hoc test suggest that the results are significantly different from those expected if the null hypothesis were correct. The post-hoc Tukey-Kramer test adjusts for multiple comparisons.*

*20.3. **More information:** P-values, confidence intervals, and effect sizes are standard means for making an inference, and any level is acceptable, though some criteria must be specified in this or previous fields. Bayesian analyses should specify a Bayes factor or a credible interval. If you are selecting models, then how will you determine the relative quality of each? In regards to multiple comparisons, this is a question with few "wrong" answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.*

INFERENCE CRITERIA:

Criterion for a successful replication attempt for the SCORE project is a statistically significant effect (alpha = .05, two tailed) in the same pattern as the original study. For this replication attempt, this criterion is met by a statistically significant interaction of stereotype activation (stereotype not activated, stereotype activated by gender-question, stereotype activated by gender difference statement)) x gender (male or female) in the ANCOVA.

---

21. Data exclusion
    - *21.1. How will you determine what data or samples, if any, to exclude from your analyses? How will outliers be handled? Will you use any awareness check?*
    - *21.2.* ***Example****: No checks will be performed to determine eligibility for inclusion besides verification that each subject answered each of the three tastiness indices. Outliers will be included in the analysis.*
    - *21.3.* ***More information****: Any rule for excluding a particular set of data is acceptable. One may describe rules for excluding a participant or for identifying outlier data.*

DATA EXCLUSIONS:

The original authors indicated in page 42/43: "*In keeping with Reips (2002) recommendations for online research, we included only those respondents in the analysis who had completed the survey without interruption*." And in page 46: "*To ensure the quality of our data we excluded participants from our sample, who indicated that they had cheated*." As in the original study, we propose to exclude participants who indicate they have cheated or used help in the political knowledge test. We also propose to exclude participants who interrupt their experimental session or who fail attention questions. Participants who initiate the survey after quotas implemented for field of study/work have been met will be redirected to the end of the survey and excluded from the sample. Likewise, participants who do not express consent to participate or state they will not answer questions openly and truthfully will be redirected to the end of the survey and excluded from the sample.

In addition, as mentioned before, we will not include in the sample participants who (a) are not of American nationality, (b) younger than 18 years old, (c) who do not identify as female or male, (d) not currently studying/working

From the paper: “To ensure the quality of our data we excluded participants from our sample, who indicated that they had cheated (n = 9). Also, we included only participants who were not registered as having interrupted their experimental session.” (p. 46)

---

22. Missing data
    - *22.1. How will you deal with incomplete or missing data?*
    - *22.2.* ***Example****: If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.*
    - *22.3.* ***More information****: Any relevant explanation is acceptable. As a final reminder, remember that the final analysis must follow the specified plan, and deviations must be either strongly justified or included as a separate, exploratory analysis.*

MISSING DATA:

We plan to follow the same procedure as in the original paper. As we understood, there weren't missing data in their study, except for the political scale which had "I don't know" option, which arguably can be considered missing data. In other words, since the political knowledge scale measures one's political sophistication, the response 'I don't know" is not missing data per se, but an indication of the participant's knowledge. To our knowledge all other questions required an answer from the participant.

23. Exploratory analysis (optional)

23.1. If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.

*23.2. **Example**: We expect that certain demographic traits may be related to taste preferences. Therefore, we will look for relationships between demographic variables (age, gender, income, and marital status) and the primary outcome measures of taste preferences.*

EXPLORATORY ANALYSIS:

We will attempt to replicate the original study results and their statistical analyses (2x2x3 ANCOVA) only. Exploratory analyses are not planned.

Other

———

24. Other (required)

24.1. If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other context that will be helpful for future readers would be appropriate here.

OTHER:

———

25. Deviations from the original study

25.1. Please describe any known deviations from the original study. Include discussion of whether and how these deviations might impact the results of this replication/reproduction attempt and its comparison to the original study.

There are two deviations from the original study, one is the sample type (students vs. adults) and another measurement of field of study (original studying politics or psychology; replication: studying **and** working). This is because the target sample for the replication is considerably high to attain at our own universities or with the survey companies we checked. For this reason, we extend the manipulation to field of study and work, so our sample also includes non-students. We ask participants to indicate which subjects are related to their studies/work. We ask this question in the beginning of the survey so we are able to have quotas implemented and later redirect participants to the end of the study in case they belong to a group for which the quota has already been filled. We provide participants with an extensive list of subjects to choose from to avoid that this question makes participants aware of the manipulation.

An important aspect of replications is whether departures from the original yields sufficient reason to question the legitimacy of a possible disagreeing result. In this specific case, we believe that the adult sample and expansion criteria to include field of work has the potential to either limit or expand the generalizability of the original author's claim. If replicated, it suggests the finding may also apply to adults more generally (it is suggestive evidence only as a nationally representative sample, ideally with a random sampling method would be necessary for stronger claims of generalizability) or, instead, if not replicated, it would suggest that the author's claims may apply to university students only, or students at Fern University Hagen).