Replication of a Research Claim from Lindqvist & Östling (2010),
from American Political Science Review

Replication Team: Levi Boxell

Research Scientist: Andrew Tyner

Action Editor: Kevin Esterling

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

Reviewer #1: [Nicholas Weller]

Reviewer #2: Landon Schnabel

Reviewer #3: [NAME]

Review Period: Feb. 27 - March 3
“Re-review”: August 17 - August 24

View-only links to: Original Paper, Original Materials

Replication of a Research Claim from Lindqvist & Östling (2010), from American Political Science Review

SCORE report LINDQVIST_AmPoliSciRev_2010_OeGv_y050

Sourcing notes for this preregistration are here.

Welcome to the replication/reproduction team! You can get started with your preregistration by clicking here.

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 the research process, including peer review. This includes privacy about which studies have been selected for replication and all aspects of the discussion about these replication designs.

Table of contents

Sourcing Notes

The research scientists write a short set of constraints or notes about each study that are designed to help match the replication/reproduction to a team that will perform it. They are maintained as part of the project record.

- SCORE Research Scientists insert sourcing notes for Chris here.

Instructions for Preregistration Reviewers

Your role is to review preregistered research designs for clarity, completeness, and quality. For the purposes of a SCORE replication, a preregistration is high quality if it generates a protocol that is a good faith attempt to replicate the original finding. In sum, focus on whether differences in original versus replication protocols are substantively anticipated to matter for claims in the original paper (and the broader field), and be biased against spending time on speculative concerns that do not have an evidence base.

Please start by reading the Reviewer Criteria checklist. For each section, you should evaluate whether the description is complete, whether deviations from the original study (or additions, if the information was not available) are documented, and whether, all told, the decisions are consistent with a good faith replication. You don't need to fill out a copy of the checklist, but should use it as a guide to the types of information that should be present in a finished preregistration - remember that not all items will apply to all projects. At the very end of the pregistration is a final reviewer checklist where you can give your evaluation of the preregistration as a whole. When you are completely finished reviewing, please comment 'DONE' on your name on the first page of this document.

These replications are intended to be robust, high quality studies, and in some cases, this will involve deviation from the original study. For instance, all SCORE projects are preregistered, and all use sample sizes that are based on a formal power analysis, whether or not these were the case in the original study. Labs may choose to include additional 'best practices' that may not have been present in the original study, in addition to other necessary differences between the original and replication study. These are allowable, so long as they remain a good faith replication of the original finding. We are collecting a list of best practices which are the kinds of steps that labs are encouraged to take (and that you can recommend!) to increase the robustness and replicability/reproducibility of their work. Keep in mind that not all projects need to (or can) include all of these practices.

The preregistration for this paper begins here. You can also reference this list of Frequently Asked Questions for preregistration reviewers.

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 the research process, including peer review. This includes privacy about which studies have been selected for replication and all aspects of the discussion about these replication designs.

Instructions for Replication/Reproduction Team

General information about preregistration is available at https://cos.io/prereg. Every section should have a response from you; in the case that a section truly does not apply (e.g., “Manipulated Variables” do not exist in an observational study), you can respond with “N/A”. Some sections are indicated as multiple choice, and we ask that you bold your response(s) if one has not already been selected. All other sections are open-ended.

A Research Scientist from the Center for Open Science has provided foundational information for your SCORE project from the original paper and, where possible, additional feedback and materials from the original author(s). **This information should not be considered a complete response to a section unless otherwise noted.**

The preregistration for your replication or reproduction should provide as detailed a plan as possible of what **you** will be doing, not just describe what was done in the original paper. That plan should be written in the future tense, and reference the original paper and any original materials or correspondence with the original author as necessary to provide context or justification of any decisions made for your protocol.

You are encouraged to look over the Reviewer Criteria that will be used to evaluate your preregistration. For each question, your response should include a complete **description** (state what you will do, in enough detail that others could implement your plan), list of **deviations** (clearly state anything you added, omitted, or changed from the original study), and **rationale** (justify why your decisions are consistent with a good-faith replication of the original claim).

Materials developed for a replication do not need to be included directly within the preregistration document. Instead, please upload any materials (such as surveys, audio/visual stimuli, instructions for coders/confederates, etc.) to the “Methods and Materials” component of the OSF project for your SCORE protocol. If you create a codebook/data dictionary or dummy dataset for review, please upload these to the “Data” component. You should also upload your data cleaning and data analysis scripts to the “Analysis” component. Although we encourage you to prepare these scripts in time for the external review of your preregistration, you are not required to do so. However, you will have to provide such a script before beginning data collection for SCORE, so we strongly encourage developing it as early as possible. Any files you upload for your project should be directly referenced within the document by filename so it is clear what is being used for your replication and where reviewers can find it. (For instance: *“Each participant will see 16 of the 64 cat pictures. All stimuli are uploaded to the Methods & Materials component, as cat1.jpg, cat2.jpg...cat64.jpg.”*)

You can reach out to the SCORE Project Coordinators for additional guidance at scorecoordinator@cos.io, and you can also reference this list of Frequently Asked Questions.

Preregistration for SCORE

Study Information

1. Title (provided by SCORE)

RR TEAM INSTRUCTIONS: *This has been determined by SCORE.*

Replication of a research claim from Lindqvist & Östling (2010) in American Political Science Review

2. Authors and affiliations

RR TEAM INSTRUCTIONS: *Fill in the names and affiliations of your team below.*

Levi Boxell[1]

1 Stanford University

3. Description of study (provided by SCORE).

RR TEAM INSTRUCTIONS: *This description has been provided by SCORE. Please review and make a SCORE project coordinator aware of any edits, additions, and corrections you would suggest to the paragraph. You are free to add additional descriptions of your project in a separate paragraph.*

The claim selected for replication from Lindqvist & Östling (2010) is that the correlation between polarization and government size is significantly stronger among democratic countries; the association between polarization and government size in strong democracies is the specific part of the finding selected for the SCORE program. This reflects the following statement from the paper's abstract: "Political polarization is strongly associated with smaller government in democratic countries, but there is no relationship between polarization and the size of government in undemocratic countries." The analysis relies on a regression of government size (general government consumption as a fraction of total consumption, averaged from 2003-2005) on the standard deviation of four World Values Survey items measuring various economic aspects of left and right on a 1-10 scale, in four separate models. The authors estimate a regression for each of these measures in the paper; the SCORE program has selected the 1-10 scale where 1 means complete agreement that “Private ownership of business should be increased” and 10 means complete agreement that “Government ownership of business and

industry should be increased." In the 'long' specification chosen for the SCORE program, the authors control for the mean value of responses, along with geographic and colonial controls, and an additional set of potentially endogenous control variables. The authors split their sample such that countries with a Polity IV democracy score of 9 to 10 are strong democracies, while countries with scores 0 to 8 are weak democracies. The evidence selected for the SCORE program is the regression estimated on the strong democratic sample using the measure of polarization specified above (results are significant using three of the four measures in the paper with this specification). When the sample is restricted to strong democracies, the estimated effect of polarization on government consumption is statistically significant (coefficient in the 'long' specification for the "Private" measure of polarization = -18.73, heteroscedasticity robust SE = 4.79, p = .01).

4. Hypotheses (provided by SCORE with possible RR team additions)

RR TEAM INSTRUCTIONS: *The focal test for SCORE is indicated as H*. If you will test additional hypotheses (or use alternate analyses) that help you to evaluate the claim your replication/reproduction is testing, number them H1, H2, H3 etc. (You can place H* in the list wherever makes sense). Please make sure that any additional hypotheses are logical deductions/operationalizations of the selected SCORE claim or are necessary to properly interpret the focal H* hypothesis. Research that is outside this scope should be described in a separate preregistration.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Are the listed hypotheses specific, concise, clearly testable, and specified at the level of operationalized variables?*
- *Are hypotheses identified as directional or non-directional, and, if applicable, have the direction of hypotheses been stated? (Example: "Customers' mean choice satisfaction will be higher in the CvSS architecture condition than in the standard attribute-by-attribute architecture condition.")*
- *Does the list of hypotheses/tests indicate whether additional hypotheses are taken from the original study or modified/added by the team?*

H* (SCORE focal test): Among strong democracies (countries with a Polity IV score of 9 or greater), polarization (as measured by the SD of the "private ownership" responses) will be negatively associated with the size of government.

Design Plan

5. Study type (provided by SCORE)

NOTE: *The study type that has been selected for you appears in bold; please do not change it.*

- 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.
- **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.**
- Meta-Analysis - A systematic review of published studies.
- Other

6. Blinding (multiple choice question)

RR TEAM INSTRUCTIONS: *Select any/all of the below that apply for your study by bolding them. You will give a longer description in the next question.*

- **No blinding is involved in this study.**
- For studies that involve human subjects, they will not know the treatment group to which they have been assigned.
- 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”)
- Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

[QUESTION 6 - BOLD YOUR RESPONSE ABOVE]

7. Blinding (free response)

RR TEAM INSTRUCTIONS: *Please describe the blinding procedures for your study here, including enough detail to allow the reviewers to evaluate your plan. If the details of a blinding procedure are closely tied to the experimental protocol, you can refer to longer descriptions (e.g. in your Data Collection response) so long as the information is available somewhere in the preregistration.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on blinding of both participants and study personnel?*

- *Does the preregistration comment on blinding of both hypotheses and condition assignment?*
- *If the original materials do not provide substantial detail on the blinding procedures, is it clear what additions the replication is making?*

N/A

8. Study Design

RR TEAM INSTRUCTIONS: *In this section, state your study design. Depending on the type of study you are conducting, this may be very brief (i.e. listing the factors and how they are manipulated, such as "2 (Color: Red/Blue) x 2 (Height:Tall/Short), between subjects"), or may be much longer. For instance, observational studies may involve more precise specification of the population and sampling strategy, or discussion of inferences involving assumptions about causal effects. Examples of study design include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Typical study designs for observation studies include cohort, cross sectional, and case-control studies.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify the unit of analysis?*
- *Does the preregistration describe how many treatment conditions will be used in the study, and how many conditions participants will be exposed to?*
- *Does the preregistration provide sufficient detail about how the study design deviates from or is congruent with the study design employed in the original study?*

Following the original study's design, a cross-sectional, country-level OLS regression of government expenditures on polarization along with the control variables will be used. Heteroskedastic robust standard errors will be used.

The original study used World Values Survey (WVS) and control variables from circa 2000 (WVS wave 4, supplemented with wave 3 when wave 4 was unavailable for a given country). The replication will use more recent WVS data from wave 6 (2010-2014) [and waves 5 (2005-2009) and 7 (2017-2020) when wave 6 is unavailable] to test the hypothesis.

The sample will be the set of countries available in wave 6 [and waves 5 and 7 when wave 6 is unavailable] of the WVS that have non-missing control variables.

9. Randomization (free response)

RR TEAM INSTRUCTIONS: *If you are doing a randomized study, state how will you randomize, and at what level. If you will not randomize some factors in your study design, please draw a clear distinction regarding which factors are randomized, and how other factors are distributed or determined across units of analysis.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration describe the level at which randomization takes place? (Examples include randomization within subject, blocked by condition, by cluster, etc.)*
- *Does the preregistration describe the method of randomization that is used? (Examples include simple, block, stratified, and adaptive covariate randomization, etc.)*
- *Does the preregistration describe how the randomization is implemented? (Examples include Kish grid, random number table, statistical software package, etc.)*

N/A

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 (multiple choice question, provided by SCORE)

1.1.1. Registration prior to creation of data
1.1.2. Registration prior to any human observation of the data
1.1.3. Registration prior to accessing the data
1.1.4. Registration prior to analysis of the data
1.1.5. Registration following analysis of the data

11. Explanation of existing data (provided by SCORE)

NOTE: *For a replication, this question refers to the data from the replication itself, not the original study. Even if we have access to the data from the original study, that is **not** the data that will be used for the replication of the claim and does not need to be included in this question.*

12. Data collection procedures

RR TEAM INSTRUCTIONS: *Please describe the process by which you will collect your data. If you are using human subjects, this should include how you will identify 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, in addition to the experimental/observational protocol itself. For studies that don't 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.*

Where details are described in other questions (e.g. study design, blinding) you can refer to those questions, so long as the complete description is provided somewhere. You are strongly encouraged to supplement your description here with materials (which might include stimuli, survey instruments, code for running data collection software, instructions for experimenters) uploaded to your OSF project, in the "Materials and Methods" component. Please use the specific file names when referencing them in your description.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration describe the target population, and how members of the target population are sampled for the study?*
- *Does the protocol describe in detail any materials that will be presented to participants, including variation or structure in those materials relevant to the experimental design? Are these materials available for review?*

The original study used WVS (World Values Survey) and control variables from circa 2000 (WVS waves 3 and 4). The dependent variable was government consumption averaged over the years 2003 to 2005.

The replication will use more recent WVS data from wave 6 (2010-2014) [and waves 5 (2005-2009) and 7 (2017-2020) when wave 6 is unavailable] to construct the measure of polarization (SD of responses to the question on government ownership) which is the main independent variable of interest.

The dependent variable (government consumption as a fraction of total consumption) will be taken from the World Bank (https://data.worldbank.org/indicator/NE.CON.GOVT.ZS). For each county, observations will be averaged over the available data points between 2010-2014 (inclusive).

The sample will be the set of countries available in wave 6 (and waves 5 and 7 when wave 6 is unavailable) of the WVS that

(a) have non-missing control variables and

(b) Have democracy scores of 9 or 10 to test H* as defined by the Polity IV (https://www.systemicpeace.org/polityproject.html) project as of 2010. (Original uses value from 2000).

(i) Note that the original self-classified 5 countries with missing democracy scores. No self-classification will be done with this replication. The original lists 33 countries as "strong democracies" (score of 9 or 10) in 2000 that have corresponding WVS data.

(ii) The Polity IV dataset lists 54 countries as "strong democracies" in 2010. Of these, 19 are in WVS wave 6. Another 11 are in WVS wave 5. Giving a total of 30 countries with a WVS wave and strong democracy score for waves 5 and 6.

Since the original paper uses observations as of 2000 for most control variables (in accordance with their polarization measures primarily coming from wave 4, 1999-2004), the control variables specified below will use observations from 2010 for most control variables (in accordance with the polarization measure primarily coming from wave 6,

2010-2014). Following the original, the choice of year for control variables will not vary with the specific WVS wave used.

The list and source of control variables are:

1. **Mean of responses to the question on government ownership**
    a. **From wave 6 (or waves 5 and 7 when wave 6 is unavailable) of the WVS**
2. **Three regional dummies (Africa, South + East Asia, Latin + South American + Caribbean).**
    a. **Defined geographically. Using Persson and Tabellini (2003) + own classification.**
3. **Three colonial variables (British, Spanish, Other) weighted by years of independence.**
    a. **Takes on value of (250 - t)/250 for countries of origin X where t indicates the years of independence and value of zero for other countries.**
    b. **Using Persson and Tabellini (2003) + own classification.**
4. **Logarithm of GDP per capita from 2010 (current US$)**
    a. **https://data.worldbank.org/indicator/NY.GDP.PCAP.CD**
    b. **Original uses data from 2000**
5. **Sum of exports and imports as share of GDP from 2010**
    a. **Imports https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS**
    b. **Exports https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS**
    c. **Original uses data from 2000**
6. **Proportion of population between 15 and 64 from 2010**
    a. **https://data.worldbank.org/indicator/SP.POP.1564.TO.ZS**
    b. **Original uses data from 2000**
7. **Proportion of population above 65 from 2010**
    a. **https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS**
    b. **Original uses data from 2000**
8. **Indicator for country having a federal political structure**
    a. **Taken from Adsera, Boix, and Payne (2003) and Persson and Tabellini (2003), and own classification (if necessary).**
9. **Indicator for OECD membership before 1993 with Turkey excluded.**
    a. **Taken from Persson and Tabellini (2003) and OECD.**

References

1. **Persson, T., and G. Tabellini. 2003. The Economic Effects of Constitutions. Cambridge, MA: MIT Press**
2. **Adsera, A., C. Boix, and M. Payne. 2003. "Are You Being Served? ` Political Accountability and Quality of Government." Journal of Law, Economics, and Organization 19 (2): 445–90.**

***Please see https://osf.io/jyf36/* for the replication data build.**

13. Sample size

RR TEAM INSTRUCTIONS: *The analytic sample sizes below come from the SCORE power analysis (see next question). These sample sizes do not account for participant attrition, data exclusions, or otherwise missing data. Your actual recruited sample will likely need to be larger than these analytic sample sizes in order to attempt to arrive at a sufficiently powered analysis. It is at your discretion to propose a sampling approach and rationale to address this difference in target **recruited** vs target **analytic** sample sizes - you can use the suggested language below or frame it in your own words.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the description of the analytic sample size (and target sample size for recruitment, if this differs) include targets for both first round and second round data collection?*
- *If there is more than one possible 'sample size' that could be referred to (cell vs. total size, cluster levels, whole design vs. subset for focal analysis), is the distinction made clear?*

Following the standard SCORE procedure, the initial target analytic sample size would be 29 strong democracies. The WVS surveys 5 and 6 contain 27 strong democracies with non-missing data, all of which will be used to test the hypothesis.

After initially testing the hypothesis using only waves 5 and 6 of the WVS, we will also data from wave 7 of the WVS to achieve the targeted 29 strong democracies when it receives IRB ethics approval.[1]

14. Sample size rationale

NOTE: *Power calculations for SCORE protocols are performed by either a Research Scientist at the Center for Open Science or by one of our consultants. In some cases, the power calculation will not yet be done for your protocol by the time you begin work on it; if you urgently require a defined sample size in order to submit your IRB application or otherwise make progress on your protocol, please contact the Project Coordinators (scorecoordinator@cos.io) so we can prioritize your protocol. Otherwise, please be patient as we complete these calculations and we will notify you when the target sample size has been defined.*

[1] In the standard SCORE procedure, if a statistically significant effect is not observed after the first round of data collection, a second round would begin. The second round of data collection wouldill sample an additional 37 strong democracies for a target pooled analytic sample of 66 strong democracies**.** Given the fact that WVS surveys 5-7 only contain 29 strong democracies with non-missing data, the second stage is infeasible.

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 collection, achieves 90% power to detect 50% of the original effect size. The power analysis for this study is found here.

15. Stopping rule

RR TEAM INSTRUCTIONS: *The first paragraph of this response refers to the two-stage data collection strategy for SCORE. Beyond this, your data collection procedures may not give you full control over your exact sample size; specify here how you will decide when to terminate your data collection. You will describe the specifics of what data is excluded in question 22, but please describe here how you will determine when to stop collecting new data, aiming to meet your analytic sample size.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *If the stopping rule is based on an estimated fall-off rate (e.g. attrition rate), is that rate justified? Does the plan specify how you will proceed if the resulting sample size is somewhat over or under the target?*
- *If the study includes post-collection data exclusion (e.g. participants are excluded who fail manipulation checks; analytic sample consists of all who fall below 1SD on some measure), does the stopping rule allow you to track inclusion without seeing the critical results?*
- *Does the preregistration make clear whether the plan for finishing data collection follows or deviates from the original study? If the original study is silent on an explicit stopping rule, is this made clear?*

The planned sample size is 29 strong democracies. After achieving that sample, planned analyses will be run. If a significant effect in the hypothesized direction is found, sampling stops. If that significant effect is not found, a second round of data collection will collect data from 37 additional strong democracies, for a pooled sample of 66 strong democracies. Sampling will stop after the second round of data collection regardless of a significant effect.

Given the fact that WVS surveys 5-7only contain 29 strong democracies with non-missing data, the second stage is infeasible.

Variables

RR TEAM INSTRUCTIONS: *The preregistration form divides variables across three questions: manipulated variables, measured variables, and indices (i.e. analytic variables derived from raw variables). Transformed variables (e.g. reaction time → log reaction time) can be defined here as well; you will discuss how those transformations are calculated in the analysis section.*

Across these questions, you should define all variables that will later be used during your analysis (including data preparation/processing). You can describe all variables in the preregistration and/or summarize and link to a data dictionary (codebook) in your repository to answer these questions; please make sure to indicate which variables are manipulated and which are measured.

If you will share data from your replication, this is also the place to state whether any variables will be removed prior to sharing the dataset (e.g. to reduce risk of participant identification or comply with copyright restrictions on scale items.)

16. Manipulated variables

RR TEAM INSTRUCTIONS: *Describe all variables you plan to manipulate and the levels or treatment arms of each variable (not applicable to an observational study). For any experimental manipulation, you should give a precise definition of each manipulated variable, e.g. "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. You can also refer to your data collection protocol if variable levels are more completely described there.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *For each of these variables, does the preregistration describe how each condition will be manipulated?*
- *Does the preregistration comment on the role of each manipulated variable in the focal analyses (e.g., independent variable, moderator, etc)?*
- *Does the preregistration describe any changes from the original study in procedure, context, or instruments used for these manipulated variables (e.g., sound condition played over headphones instead of speakers)?*

N/A

17. Measured variables

RR TEAM INSTRUCTIONS: *Describe each variable you will measure, including outcome measures, as well as any predictors, covariates, or descriptive information that you will measure. 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.'*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on the role of each measured variable in the focal analyses (e.g., inclusion criteria, dependent variables, control variables, etc)?*
- *If the study will measure variables which will not be involved in the focal analysis, are these variables disclosed?*
- *Does the preregistration describe any changes in procedure, context, or instruments used for these measured variables (e.g., extraversion measured with EPI vs MIES)?*

See 12.

18. Indices

RR TEAM INSTRUCTIONS: *If any of the measured variables described in Section 17 are going to be combined into a composite measure (including simply a mean), describe what measures you will use and how they will be combined. Include either a formula or a precise description of your method. If you 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.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify each of the composite measures (e.g. mean scores, factor scores) that are needed for the focal analysis, and which of the measured variables in Section 17 are used in each one (e.g. the happiness, joy, and satisfaction items will be used to create the 'positive feelings' measure)?*
- *Does the preregistration provide a detailed description of the methodology or a precise formula that will be used to construct each composite measure?*

See 12.

Analysis Plan

19. Statistical models

RR TEAM INSTRUCTIONS: *This section should describe in detail the analysis that will be performed to replicate the focal result. This analysis must align as closely as possible with the original study's analysis, even if you have identified limitations in the original study. The level of detail should allow anyone to reproduce your analyses from your description below. Examples of what should be specified: the model; each variable; adjustments made to the standard errors and to case weighting; additional analyses that are required to set up the focal analysis ; the software used.*

Beyond the replication of the focal analysis from the original study, it is at your discretion to test the claim using other analytic approaches as a check of the robustness of the claim. The original test should be listed first and be clearly distinguished from any other tests. If you are testing additional confirmatory hypotheses, describe them in the same order as you numbered them in the "Hypotheses" section above and make clear reference to the specific hypothesis being tested for each.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify which statistical model will be used to provide the 'focal evidence' for the SCORE test (e.g. a regression coefficient in a larger multiple regression model), and does it correspond closely to the model and evidence from the original study?*
- *Does the preregistration describe each variable that will be included in the focal analysis, and what role each variable has (e.g. dependent variable, independent variable)?*
- *Does the preregistration include a detailed specification of the focal analysis, including interactions, lagged terms, controls, etc.?*

For the purposes of SCORE, to test **H*** in line with the original paper, [I/we] will use...

an OLS regression of government consumption as a proportion of total consumption on the standard deviation of responses to the private ownership question in the WVS along with the control variables outlined in 12. Heteroskedastic robust standard errors will be used.

Please see https://osf.io/kt7c4/ for the analysis code that will be used.

20. Transformations

RR TEAM INSTRUCTIONS: *This section should describe how any of the measured variables or composite measures mentioned above will be transformed prior to the analyses listed in Section 19. These are adjustments made to variables **after** measurement or measure creation, and might include centering, logging, lagging, rescaling etc. Please provide enough detail such that anyone else could reproduce the transformations based on the description below.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify which of the measured variables or composite measures will need to be transformed prior to the focal analysis?*
- *For each variable needing transformation, does the preregistration adequately describe the transformations, including any centering, logging, lagging, recoding, or implementation of a coding scheme for categorical variables?*
- *For any categorical predictors that are included in a regression, does the preregistration indicate how those variables will be coded (e.g. dummy coding, summation coding, etc.) and what the reference category will be?*

See 12.

21. Inference criteria

RR TEAM INSTRUCTIONS: *This section describes the precise criteria that will be used to assess whether the hypotheses listed above were confirmed by the analyses in Section 19. The default language below only applies to the test of the SCORE claim, **H***. It is at your discretion to describe the inferential criteria you will use for any additional analyses. They need not rely on p-values and/or the same alpha level we have specified for **H***.*

If the additional analyses will use multiple comparisons, the inference criteria 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.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *For each hypothesis listed in the Hypotheses section above, does the preregistration clearly describe the inference criteria necessary to call the replication attempt for that hypothesis test successful, specified at the level of the operationalized variables?*
- *Example:"...For this replication attempt, this criteria is met by a statistically significant difference in customers' mean choice satisfaction in the CvSS architecture condition compared to the standard attribute-by-attribute architecture condition, with the mean*

higher in the CvSS architecture condition than in the standard attribute-by-attribute architecture condition."

Criteria 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 on the focal hypothesis test (**H***). For this study, this criteria is met by a negative and significant coefficient on the term for polarization in the focal regression model.

Following the original, heteroskedastic robust standard errors will be used. Specifically, HC1 standard errors computed using vcovHC in R, which corresponds to Stata's default "cluster" option.

22. Data exclusion

RR TEAM INSTRUCTIONS: *The section below should describe the rules you will follow to exclude collected cases from the analyses described in Section 19. Note that this refers to exclusions **after** data collection; exclusion criteria that prevent a case from entering your recruitment sample should be described in earlier sections. Please be as detailed as possible in describing the rules you will follow (e.g. What is the specific definition of outliers you will use? Exactly how many attention checks does a participant need to fail before their removal from the analytic sample?).*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on whether any cases recruited for the study sample will be excluded prior to data analysis?*
- *If yes, does the preregistration provided detailed instructions on how the exclusions will be performed (e.g. Is the definition of outlier provided? Is the number of attention checks failed before a participant is excluded specified?)*

Only countries that have missing data or those that do not have the full 10 point scale to the WVS question will be excluded.

23. Missing data

RR TEAM INSTRUCTIONS: *The section below should describe how missing or incomplete data will be handled. Please be as detailed as possible in describing the exact procedures you will follow (e.g. last value carried forward; mean imputation) and any software required (e.g. We will use Amelia II in R to perform the imputation).*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on how missing or incomplete data will be addressed (e.g. casewise removal, missing data imputation)?*

- *If applicable, does the preregistration specify how many missing variables will lead to a case's removal (e.g. If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.)?*
- *If applicable, does the preregistration describe how missing data imputation will be performed, including relevant software?*

Countries with missing data will be excluded.

24. Exploratory analysis (Optional)

RR TEAM INSTRUCTIONS: *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. If any exploratory analyses involve additions to the data collection procedure beyond what was performed in the original study (e.g. additional items on the survey; running another condition in the experiment), please describe them below.*

N/A

25. Other

RR TEAM INSTRUCTIONS: *This section serves two purposes. First, please use this section to discuss any features of your replication plan that are not discussed elsewhere. Literature cited, disclosures of any related work such as replications or work that uses the same data, plans to make your data and materials public, or other context that will be helpful for future readers would be appropriate here. Second, please also re-surface any major deviations from earlier in the preregistration that you expect a reasonable reviewer could flag for concern. Give a summary of these deviations, focusing on larger changes and any possible challenges for comparing the results of the original and replication study.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration reference other sections of the preregistration where substantial deviations from the original study have been described (including deviations due to differences in location or time compared to the original study)?*
- *Does the preregistration comment on plans to make the data and materials from the replication study public?*

Data and code for the replication will be made publically available.

Final review checklist

REVIEWER INSTRUCTIONS: *For the following questions, reviewers please indicate whether you can 'sign off' on the following items by adding a comment. You can update this response as the lab moves through revisions during the review period!*

- Included with this pre-registration are the specific materials needed to conduct the replication (surveys, stimuli, etc). If not,
    - Have the pre-registration authors detailed when these materials will be made available prior to final registration?
    - Can you evaluate whether this preregistration represents a good-faith replication of the original study without seeing these materials?
- Included with this pre-registration are the specific analytic scripts/code/syntax that will be used for the final analysis. If not,
    - Have the pre-registration authors detailed when these analyses will be made available prior to final registration?
    - Can you evaluate whether this preregistration represents a good-faith replication of the original study without seeing these materials?
- I have reviewed all sections of this pre-registration, and I believe it represents a good-faith replication attempt of the original focal claim.