Replication of a Research Claim from McLaren (2012),
from World Politics

Replication Team: Marta Kołczyńska

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: Bert Bakker

Reviewer #2: [NAME]

Reviewer #3: [NAME]

Review Period: June 15 - June 22

View-only link to: Original Paper

Replication of a Research Claim from McLaren (2012), from World Politics

SCORE report McLaren_WorldPolitics_2012_wRvv_1634

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 McLaren (2012), in World Politics

2. Authors and affiliations

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

Marta Kołczyńska[1]

1 Institute of Political Studies of the Polish Academy of Sciences

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 McLaren (2012) is that individuals expressing most concern about the impact of immigration on the national community will be most distrusting of politicians and political institutions (Proposition 1). This reflects the following statement from the paper's abstract: "The findings indicate that even after controlling for other predictors of trust in the political system, concerns about the effect of immigration on the national community have an impact on trust in politics." The claim is tested with multivariate analyses conducted using HLM on four rounds of the ESS [European Social Survey]. The author uses multilevel modeling with a three-level model with the individual at level 1, variables that are measured at the country level and that vary across the four rounds of the ESS (country-round) at level 2, and variables measured at the country level that do not vary across the four rounds at level 3. The dependent variable, political distrust, measures how much the respondent trusts each of three institutions: the country's parliament, the legal system, and politicians; for the SCORE program, the parliament measure is used [the author finds support using each measure]. The independent variable of interest is concern about immigration (see the Parliament columns in Table 3 for

details on the model selected for the SCORE program). The results indicate that after controlling for fairly powerful predictors of distrust in politics, concern about immigration has a statistically significant effect on distrust in politics, with maximum effects of 1.7 on the 11-point measure of distrust in parliament (coefficient [b] on concern about immigration term = 0.17, SE = 0.00, p = 0.000).

Concern about immigration ranges between 0 and 10, and moving from 0 to one would be associated with a 1.7 unit change on distrust in the parliament, which also is measured on a scale from 0 to 10.

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): Individuals' concerns about immigration is positively associated with distrust in their country's parliament.

[CONTINUE QUESTION 4 RESPONSE HERE]

No additional hypotheses will be tested.

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?*

[QUESTION 7 RESPONSE HERE]

Not applicable: the original study relied on secondary analysis of available cross-national survey data, and the replication will use a different dataset of the same type.

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?*

[QUESTION 8 RESPONSE HERE]

The original study is an observational study relying on secondary analysis of cross-national repeated cross-sectional survey data from the European Social Survey (ESS), pooled from rounds 1-4 (carried out between 2002 and 2009), from 16 Western European countries: Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France,the United Kingdom, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, and Sweden. The total analyzed data subset contained 110,732 cases corresponding to individual respondents, from those 16 countries, with between 2,698 (Italy) and 10,719 (Germany) cases per country. Since not all countries were surveyed in all 4 rounds of the ESS, the total number of country-rounds was 59.

National samples in ESS aim to cover entire adult populations of the respective country, defined as individuals aged 15 or above. Probably sampling is applied and violations of the probability design, such as respondent substitutions, are not permitted. Overall, the ESS is known for the

high quality of the data and emphasis on data comparability and measurement equivalence, and is widely used in academic research.

It is worth noting that the number of cases in the original study (110,732) likely includes all cases in the data, not accounting for missing data. I suspect this based on looking at the data from ESS/1-4, and also because all three analyses in Table 3 have the same number of individual cases in the table's footnote, while given that these analyses have different dependent variables, the actual number of cases in each model is likely slightly different. Looking at the ESS data from rounds 1-4, it seems that around 25% of all cases have missing values on the household income variable, while about 10% have missing values on self-placement on the left-right scale, both of which are included in the analysis. Item non-response on other variables is smaller. The author does not mention imputation to fill the missing data. If listwise deletion was used, the final sample size was likely around 73 thousand cases. The article does discuss missing values.

The replication will consist in secondary analysis of the same survey project - ESS - but from round 5, carried out in 2010-2011. Using just one round of ESS data will reduce the sample size to 25,586, with between 1226 (Switzerland) and 2821 (Germany) cases per country. While smaller than the original study, this sample size is still very large, especially given that the tested hypothesis does not involve interaction effects.

Using pooled data from the next four consecutive rounds of ESS - round 5-8 - carried out between 2010 and 2017, was also considered. However, the choice of ESS Round 5 only seems preferable since it refers to a time before the refugee crisis, i.e. the increased inflow of refugees and migrants to Europe from the Middle East and North Africa in the second half of 2010s. Since the refugee crisis has increased the size of foreign-born populations in many Western European countries and made issues related to refugee status and immgiration more salient, using survey data from that time could potentially affect the results of the replication.

Using the most proximate round of ESS will generally minimize the changes in the context and conditions in Europe that could potentially influence the results. Given the high level of standardization of ESS across countries and rounds of the project, this choice of data minimizes the number of deviations from the original study.

Alternative data sources from surveys carried out between 2002 and 2009, as the original study, include the European Values Study or the Eurobarometer. However, these projects do not have all the individual-level variables that the original study uses as controls.

Following the original study, the replication will include only Western European countries. ESS round 5 only covered 14 Western European countries: Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, the United Kingdom, Greece, Ireland, the Netherlands, Norway, Portugal, Sweden. The smaller number of countries (14 compared to 16 in the original

study) should not make a difference, since the hypothesized effect is modeled as constant across countries.

The total number of cases equals 25,586 (after excluding members of ethnic minorities and non-citizens, as in the original study), but this includes records with missing values. Here, again, the variables with the highest share of missing values are household income and self-placement on the left-right scale. After eliminating all cases with missing values the dataset has 17,134 cases.

The core questionnaire used in ESS round 5 is available here:
https://www.europeansocialsurvey.org/docs/round5/fieldwork/source/ESS5_source_main_questionnaire.pdf
The documentation is here:
https://www.europeansocialsurvey.org/docs/round5/survey/ESS5_data_documentation_report_e04_2.pdf

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.)*

[QUESTION 9 RESPONSE HERE]

Not applicable.

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.*

The replication, just like the original study, will consist in secondary analysis of cross-national survey data from the European Social Survey round 5, carried out in 2010-2011.

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?*

[QUESTION 12 RESPONSE HERE]

The replication will consist in secondary analysis of available data. The subset of the data for analysis will be selected according to the same criteria as in the original study: only data from Western European countries will be included, and only citizens and respondents who do not declare being members of ethnic minorities will be analyzed.

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?*

For these calculations, the primary unit of analysis is respondents nested within countries. An estimate for the minimum required sample size is **a sample of 353**. Stage 1 and Stage 2 sample sizes were also calculated, which are **1717** and **3862** respectively.

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.*

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.

For data analytic replications in SCORE, three sample sizes are calculated:

- A minimum threshold sample size, defined as the sample size required for 50% power of 100% of the original effect
- A stage 1 sample size, defined as the sample size needed to have 90% power to detect 75% of the original effect
- A stage 2 sample size, defined as the sample size needed to have 90% power to detect 50% of the original effect

Details about how those sample sizes were calculated for this project are 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?*

Not applicable: the replication will consist in secondary data analysis. All available observations will be used in a single analysis.

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)?*

[QUESTION 16 RESPONSE HERE]

Not applicable.

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)?*

[QUESTION 17 RESPONSE HERE]

Following the original study, the replication will include characteristics of individuals (from the European Social Survey data), and characteristics of countries in the survey year (from external non-survey data sources).
The selection of survey variables was aided by a detailed appendix in the original article.

Individual-level variables

Trust in the national parliament: 11-point scale, 0 = the least distrust, 10 = the most distrust.
* The original coding was reversed to have high scores correspond to high distrust.

Concern with immigration scale: Measured on 11-point scales from 0 = the least concern with immigration to 10 = the most concern with immigration.

1. Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries?
2. Would you say that [country]'s cultural life is generally undermined or enriched by people coming to live here from other countries?
3. Is [country] made a worse or a better place to live by people coming to live here from other countries?

* The original coding was reversed to have high scores correspond to high concern with immigration.

Unhappiness: Taking all things together, how happy would you say you are? Please use this card. Extremely unhappy (0), Extremely happy (10).
* The original coding was reversed to have high scores correspond to high unhappiness.

Dissatisfaction with life: All things considered, how satisfied are you with your life as a whole nowadays? Please answer using this card. Extremely dissatisfied (0), Extremely satisfied (10).
* The original coding was reversed to have high scores correspond to high dissatisfaction.

Frequency of meeting friends: Using this card, how often do you meet socially with friends, relatives or work colleagues? Never (1), Less than once a month (2), Once a month (3), Several times a month (4), Once a week (5), Several times a week (6), Every day (7).
* The original coding reversed to have high values correspond to infrequent meeting with friends.

Interpersonal distrust:

Using this card, generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted.

Using this card, do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair? Most people would try to take advantage of me (0), Most people would try to be fair (10).

Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves? Please use this card. People mostly look out for themselves (0), People mostly try to be helpful (10).

* The original coding reversed to have high values correspond to high distrust.

Dissatisfied with country's economy: On the whole how satisfied are you with the present state of the economy in [country]"? Extremely Dissatisfied (0), Extremely satisfied (10).
* The original coding reversed to have high values correspond to high dissatisfaction.

Dissatisfied with personal income: Which of the descriptions on this card comes closest to how you feel about your household's income nowadays?" Living comfortably on present income (1), Coping on present income (2), Finding it difficult on present income (3), Finding it very difficult on present income (4).

Dissatisfied with health system: Still using this card, please say what you think overall about the state of health services in [country] nowadays? Extremely bad (0), Extremely good (10).
* The original coding reversed to have high values correspond to high dissatisfaction.

Dissatisfied with education system: Now, using this card, please say what you think overall about the state of education in [country] nowadays? Extremely bad (0), Extremely good (10).
* The original coding reversed to have high values correspond to high dissatisfaction.

Voting for winning party or candidate: Some people don't vote nowadays for one reason or another. Did you vote in the last [country] national election in [month/year]? IF "YES" ⇒ Which party did you vote for in that election?
* Whether the party the Respondent voted for won or lost the election (i.e., was part of the government following that election or not) is coded using data from the Parliaments and governments database (ParlGov, http://www.parlgov.org/) keeping in mind the fieldwork dates in the respective countries.

Voted for far-right party in last general election: Coded on the basis of the "party voted for" variable and list of far-right-wing parties from the original study's Web Appendix available at http://eprints.nottingham.ac.uk/1566/2/McLaren_Cultural_Divide_in_Europe_Web_Appendix.pdf

Left-right scale: In politics people sometimes talk of "left" and "right." Using this card, where would you place yourself on this scale, where 0 means the left and 10 means the right?

HH income: Using this card, please tell me which letter describes your household's total income, after tax and compulsory deductions, from all sources? If you don't know the exact figure, please give an estimate. Use the part of the card that you know best: weekly, monthly, or annual.
Income was standardized within each country.

Age: In what year were you born?
* Age derived from year of birth by data producers.

Education: What is the highest level of education you have achieved?
1 = ES-ISCED I, less than lower secondary;
2 = ES-ISCED II, lower secondary;
3 = ES-ISCED IIIb, lower tier upper secondary OR ES-ISCED IIIa, upper tier upper secondary;
4 = ES-ISCED IV, advanced vocational, sub-degree;
5 = ES-ISCED V1, lower tertiary education, BA level;
6 = ES-ISCED V2, higher tertiary education, >= MA level.

Female: Gender coded by interviewer, recoded to: 0=Male and 1=Female.

Country-level variables

Control variables:

Far-right party popularity: Calculated for the parties identified as far-right in the original study's Web Appendix available at http://eprints.nottingham.ac.uk/1566/2/McLaren_Cultural_Divide_in_Europe_Web_Appendix.pdf

Social protection expenditure: This is measured by the total expenditure on social protection per head of population in ecu/euro, in the year before each survey, based on data from the Eurostat: https://ec.europa.eu/eurostat/databrowser/view/tps00099/default/table?lang=en

Long-term country of immigration, post–World War II (dummy): As in the original study, Greece, Spain, Portugal, Italy, and Ireland are assigned the value of 0 for this analysis, and all other countries are assigned the value of 1.

World Bank Governance Indicators: six indicators created by the World Bank's Worldwide Governance Indicators: Voice and Accountability, Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. (https://datacatalog.worldbank.org/dataset/worldwide-governance-indicators).
The six indicators are combined into one using the mean.
WGI_csv.zip downloaded on 11.01.2020 from
https://databank.worldbank.org/reports.aspx?source=worldwide-governance-indicators#

GDP per capita: measured in the year before the survey from the OECD data, USD constant prices 2015 PPP: https://stats.oecd.org/Index.aspx?DataSetCode=PDB_LV

Unemployment: Unemployment rate in the year before the survey from the OECD data: https://data.oecd.org/unemp/unemployment-rate.htm

17.1. Data Dictionary (as in point 12e in the Existing Data Replication template)

RR TEAM INSTRUCTIONS: *Create a data dictionary following this template. Provide below a view-only link to the completed data dictionary included in the OSF project. If the Data Analyst will need to create new variables using the variables in the final replication dataset (e.g. recoding the provided education variable to be in a better format for analysis), please document below your recommendation on how the analyst should do so. Please also document any additional notes regarding the variables in the dataset that do not fit within the provided data dictionary template or the other sections above.*

Link to dictionary:
https://docs.google.com/spreadsheets/d/1p-hzj8tLb5ZCeHc6cid41tQojQgRhfxROAyS7GsG7uo/edit?usp=sharing

Link to the data cleaning code:
https://osf.io/v9dzc/?view_only=06a7760c57024488bc52139984d62f97

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?*

[QUESTION 18 RESPONSE HERE]

Individual-level:
Concern with immigration = mean of reversed scores for: imbgeco, imueclt, imwbcnt.
Interpersonal distrust = mean of reversed scores for: ppltrst, pplfair, pplhlp.

In both cases missing values are ignored, i.e. if there are two valid responses, the index takes the mean of the two. If there are no valid responses, the index is missing as well.

Country-level:
Quality of governance = mean of 6 Quality of Governance indicators: Voice and Accountability, Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption.

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...

[CONTINUE QUESTION 19 RESPONSE HERE]

Main analysis: Two-level hierarchical linear modeling with individuals nested in countries, and with "distrust in parliament" as the dependent variable, the "concern with immigration" index as the independent variable, and all the individual- and country-level control variables.

This mimics the original study design, which used data from rounds 1-4 of the European Social Survey, and a three-level model with individuals nested in country-years, nested in countries.

The replication will apply post-stratification and design weights as provided in the ESS data. The original study does not mention weights at all, but using them is recommended practice.

The main analysis will use listwise deletion, i.e. eliminating all cases with missing values on the variables used in the analysis.

Additional analyses, intended as robustness checks, will include:

1. The model described above without survey weights,
2. The model described above with imputed missing data,
3. The model described above with imputed missing data and without survey weights.

Link to the imputation code (applied to 5% of the data):
https://osf.io/um7wv/?view_only=06a7760c57024488bc52139984d62f97

Link to the analysis code - primary analysis and three auxiliary analyses as listed above - and model output for 5% of the data:
https://osf.io/cqmza/?view_only=06a7760c57024488bc52139984d62f97

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?*

[QUESTION 20 RESPONSE HERE]

Country-level variables referring to economic development, quality of governance, social expenditure, and unemployment, will be taken from the year prior to the year of the survey. Popularity of right-wing parties will be measured using results of the most recent general election.

Education will be recoded into the categories used in the original study according to the following schema:

source label	source value	recoded to
Not possible to harmonise into ES-ISCED	0	missing
ES-ISCED I , less than lower secondary	1	1
ES-ISCED II, lower secondary	2	2
ES-ISCED IIIb, lower tier upper secondary	3	3
ES-ISCED IIIa, upper tier upper secondary	4	3
ES-ISCED IV, advanced vocational, sub-degree	5	4
ES-ISCED V1, lower tertiary education, BA level	6	5
ES-ISCED V2, higher tertiary education, >= MA level	7	6
Other	55	missing

While education measured in levels is an ordinal variable, following the original study, it will be entered into the regression model as a continuous predictor (which is also common practice in analyses of survey data).

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 positive and significant coefficient on the term for concern about immigration in the focal regression model.

[CONTINUE QUESTION 21 RESPONSE HERE]

[nothing to add]

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?)*

[QUESTION 22 RESPONSE HERE]

From the source data set from the European Social Survey, the following observations will be excluded in the main analysis:

1. Respondents surveyed in countries outside of Western Europe (understood as pre-2004 European Union or European Free Trade Association),
2. Respondents who are not citizens of the country of residence,
3. Respondents who declared being part of an ethnic minority in the country of residents,
4. Cases for which any of the variables of interest has missing values (case deletion).

Additional analyses intended as robustness checks (see section 19. Statistical models) will use imputation as a way of addressing missing data.

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?*

[QUESTION 23 RESPONSE HERE]

In the main analysis, cases with any missing values on any of the variables of interest will be removed from the analysis.
As written in section 8. Study Design, the total number of cases in the data equals 25,586, but this includes records with missing values. After eliminating all cases with missing values the dataset has 17,134 cases.

Additional analyses intended as robustness checks (see section 19. Statistical models) will use imputation as a way of addressing missing data.

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.*

[no exploratory analyses will be conducted]

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?*

[QUESTION 25 RESPONSE HERE]

Disclaimer: The author of the preregistration has prior experience with analyzing data from the European Social Survey, including many of the variables used in the present replication, in particular trust in parliament and socio-demographic variables.

The primary deviation of the replication from the original study consists in using data from a different (later) round of the European Social Survey. The original study pooled data from rounds 1-4 (2002-2009) and analyzed data from 16 Western European countries. The replication will use data from round 5 (2010-2011), which included data from 14 Western European countries. The sample sizes of individual surveys (country-year) are generally similar and in the range between 1500 and 3000 respondents. In the original study there was more than one sample from each country, each from a different ESS round, i.e. a different year. In the replication there will be one sample from each country.
Since the replicated hypothesis refers to the association between concerns with immigration and distrust in the parliament net of other factors, and does not specify any pattern of change over time, I believe that an analysis of round 5 of the ESS constitutes a good-faith replication of the original study.

The analysis plan includes a main analysis and additional analyses, as described in Section 19. Statistical models. Additional analyses are meant to ensure the robustness of the results to the imputation of missing values and presence/absence of survey weights.

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.