

# Transparency in Social Science Research & Teaching

**Dr Nicole Janz**

# The Economist

OCTOBER 19TH-25TH 2013

[Economist.com](http://Economist.com)

Britain's angry white men

How to do a nuclear deal with Iran

Investment tips from Nobel economists

Junk bonds are back

The meaning of Sachin Tendulkar

# HOW SCIENCE GOES WRONG

99  
Einsteinium

# Why Do So Many Studies Fail to Replicate?

---

**Gray Matter**

By JAY VAN BAVEL MAY 27, 2016

---



Unreliable research

# Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition



22K



"I SEE a train wreck looming," warned Daniel Kahneman, an eminent psychologist, in an

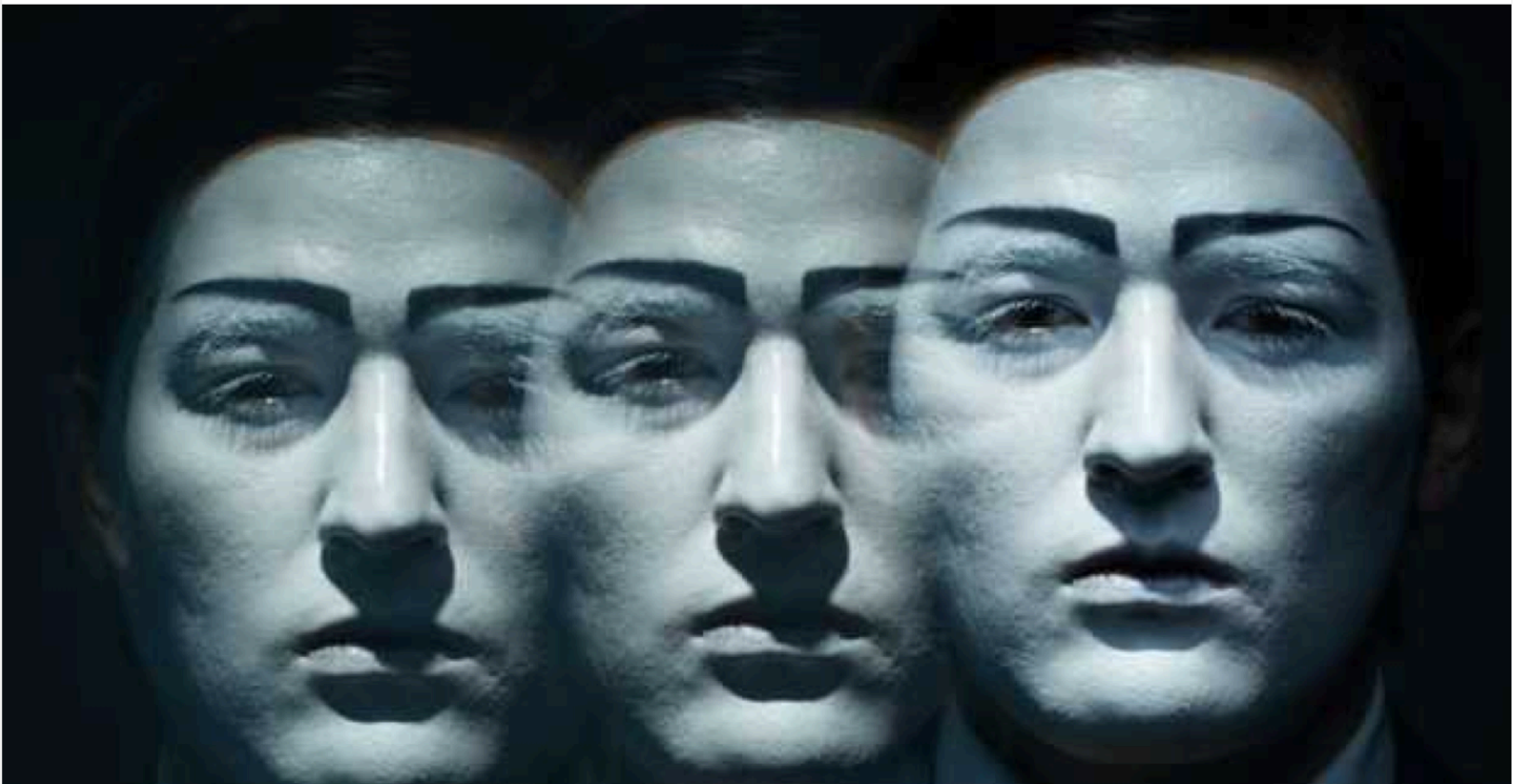


# More than half of psychology papers are not reproducible

Initiative to replicate findings of 100 prominent studies raises further questions about health of discipline

August 27, 2015

By [Paul Jump](#) Twitter: [@PaulJump](#)





# How to replicate

# What is a replication?

## Duplication

Verify research results

*exact same data set*  
*exact same methods*

## Replication

Test the robustness of the original research results

*new data*  
*new models*



# Which study should I pick?

**Relevant research  
with impact**

**Outdated  
measures**

## The perfect replication project

### Abstract

I'm the perfect replication project because I combine all these, or at least most of these, features: interesting & relevant questions, results that are accepted but have never been checked, fail to control for important variables, use out-dated measurements, make you wonder if the results apply in different contexts, I'm pointed at in "limitations" and "future research" sections of articles, I'm in an area 'ripe for replication'.

Keywords: *replication, relevant, improvement*

**Results widely  
accepted but  
never checked**

**Missing control  
variables**

# Examples of a ‘good pick’

Reinhart & Rogoff. 2010.  
“Growth in a Time of Debt.”

**Argument:** high debt is associated with lower growth

**Impact:**

- high journal (The American Economic Review)
- research was used by governments to justify austerity measures

*American Economic Review: Papers & Proceedings* 100 (May 2010): 573–578  
<http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.573>

## Growth in a Time of Debt

By CARMEN M. REINHART AND KENNETH S. ROGOFF\*

In this paper, we exploit a new multi-country historical dataset on public (government) debt to search for a systemic relationship between high public debt levels, growth and inflation.<sup>1</sup> Our main result is that whereas the link between growth and debt seems relatively weak at “normal” debt levels, median growth rates for countries with public debt over roughly 90 percent of GDP are about one percent lower than otherwise: average (mean) growth rates are several percent lower. Surprisingly, the relationship between public debt and growth is remarkably similar across emerging markets and advanced economies. This is not the case for inflation. We find no systematic relationship between high debt levels and inflation for advanced economies as a group (albeit with individual country exceptions including the United States). By contrast, in emerging market countries, high public debt levels coincide with higher inflation.

Our topic would seem to be a timely one. Public debt has been soaring in the wake of the recent global financial maelstrom, especially in the epicenter countries. This should not be surprising, given the experience of earlier severe financial crises.<sup>2</sup> Outsized deficits and epic bank bailouts may be useful in fighting a downturn, but what is the long-run macroeconomic impact,

especially against the backdrop of graying populations and rising social insurance costs? Are sharply elevated public debts ultimately a manageable policy challenge?

Our approach here is decidedly empirical, taking advantage of a broad new historical dataset on public debt (in particular, central government debt) first presented in Carmen M. Reinhart and Kenneth S. Rogoff (2008, 2009b). Prior to this dataset, it was exceedingly difficult to get more than two or three decades of public debt data even for many rich countries, and virtually impossible for most emerging markets. Our results incorporate data on 44 countries spanning about 200 years. Taken together, the data incorporate over 3,700 annual observations covering a wide range of political systems, institutions, exchange rate and monetary arrangements, and historic circumstances.

We also employ more recent data on external debt, including debt owed both by governments and by private entities. For emerging markets, we find that there exists a significantly more severe threshold for total gross external debt (public and private)—which is almost exclusively denominated in a foreign currency—than for total public debt (the domestically issued component of which is largely denominated in home currency). When gross external debt reaches 60 percent of GDP, annual growth declines by about two percent; for levels of external debt in excess of 90 percent of GDP, growth rates are roughly cut in half. We are not in a position to calculate separate total external debt thresholds (as opposed to public debt thresholds) for advanced countries. The available time-series is too recent, beginning only in 2000. We do note, however, that external debt levels in advanced countries now average nearly 200 percent of GDP, with external debt levels being particularly high across Europe.


The focus of this paper is on the longer term macroeconomic implications of much higher public and external debt. The final section, however, summarizes the historical experience of the United States in dealing with private sector

\*Reinhart: Department of Economics, 4115 Tydings Hall, University of Maryland, College Park, MD 20742 (e-mail: creinhart@umd.edu); Rogoff: Economics Department, 216 Littauer Center, Harvard University, Cambridge MA 02138-3001 (e-mail: krogoff@harvard.edu). The authors would like to thank Olivier Jeanne and Vincent R. Reinhart for helpful comments.

<sup>1</sup> In this paper “public debt” refers to gross central government debt. “Domestic public debt” is government debt issued under domestic legal jurisdiction. Public debt does not include debts carrying a government guarantee. Total gross external debt includes the external debts of all branches of government as well as private debt that is issued by domestic private entities under a foreign jurisdiction.

<sup>2</sup> Reinhart and Rogoff (2009a, b) demonstrate that the aftermath of a deep financial crisis typically involves a protracted period of macroeconomic adjustment, particularly in employment and housing prices. On average, public debt rose by more than 80 percent within three years after a crisis.

# Practical steps in a replication study

- 1 Select paper
  - 2 Access data & code
  - 3 Identify each variable
  - 4 Reproduce tables, figures
  - 5 Compare
- 
- 2-3 weeks
- 3-4 weeks

If you got to this point, you completed a **duplication**.

# Practical steps in a replication study (II)

## 6 Add value

- new data
- new variables
- new model specifications
- theoretical contributions



4-6 weeks

## 7 Compare

## 8 Get feedback from peers

## 9 Journal submission



months

**You now completed a full replication!**



# Communicating failed replications



# What replicators write



“We ... find that coding errors, selective exclusion of available data, and unconventional weighting of summary statistics lead to **serious errors**” (Herndon et al. 2013)

“If we cannot even reproduce the original results using the same publicly available data, there is **no need for further commentary.**” (Miller et al, 2001)

# How original authors often **respond**

“less realistic”, “inconsistent with the substantive literature,” and “**of limited utility**” (Mansfield, Milner, and Rosendorff 2002)



“fundamentally **flawed**”  
(Peffley, Knigge, and Hurwitz 2001)

“statistical, computational, and reporting errors that **invalidate its conclusions**” (Gerber and Green 2005:301).

# Publishing a replication study

- Good replication studies get published
- Write a **solid** paper (puzzle, relevance, hypothesis, research design, findings, discussion) – as if it was an **original** piece.
- In some fields (politics): **Don't sell it as a replication** paper



# Voting Costs and Voter Turnout in Competitive Elections

Bernard Fraga<sup>1</sup> and Eitan Hersh<sup>2,\*</sup>

Our estimation approach builds off of the methodology and data used by Gomez et al. (2007) ..., adding measures of electoral closeness in order to focus on how the randomly assigned cost (rain) has a different impact depending on the electoral environment.

same way even to rain, then serious doubt should meet claims that voters will react

# Political Regimes and International Trade: The Democratic Difference Revisited

XINYUAN DAI *University of Illinois at Urbana–Champaign*

**H**ow do domestic political institutions affect the outcomes of international trade negotiations? Specifically, are the aggregate trade barriers agreed upon by a democratic pair lower than those by a pair composed of a democracy and an autocracy? I revisit these important questions

by highlighting  
Contrary to  
democratic  
Thus, I  
level of

I revisit these important questions by highlighting some problematic aspect of the analysis by Mansfield, Milner, and Rosendorff (2000).

**H**ow do domestic political institutions affect the outcomes of international trade negotiations? Specifically, are the aggregate trade barriers agreed upon by a democratic pair lower than those by a pair composed of a democracy and an autocracy? I revisit these important questions, analyze a model in which decision makers agreed upon trade barriers, and find that the aggregate trade barriers agreed upon by a democratic pair are lower than those agreed upon by a mixed pair.

To distinguish between country A, or a legislature while an autocrat

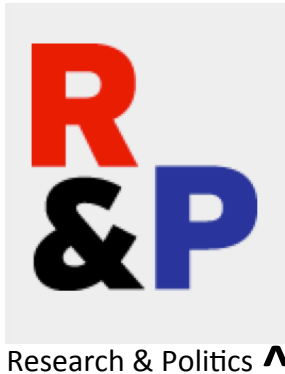
Contrary to their central conclusion, I find that whether the aggregate trade barriers are lower for a democratic pair than those for a mixed pair depends on the preferences of the decision makers involved.

TILI games played by democratic pairs, by autocratic

ideal level of trade barriers at home and abroad for

# Journals Open to Replication (selection)

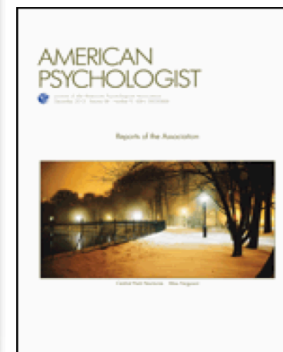
## Political Science



## Psychology



## Economics



\*

\*

+

#

- \* original study was published in the same journal
- + home of the original 'Many Labs' project
- # special issue dedicated to replications (March 2015)
- ^ this journal invites replication studies

# Replications by Early Career Researchers



POLITICAL ECONOMY RESEARCH INSTITUTE

## Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

Thomas Herndon, Michael Ash and Robert Pollin

Psychophysiology, 52 (2015), 159–166. Wiley Periodicals, Inc. Printed in the USA. Copyright © 2015 The Authors. Psychophysiology published by Wiley Periodicals, Inc. on behalf of Society for Psychophysiological Research. DOI: 10.1111/psyp.12136

### Near-wins and near-losses in gambling: A behavioral and facial EMG study

YIN WU,<sup>1</sup> ERIC VAN DIJK,<sup>2</sup> and LIJKE CLARK<sup>1\*</sup>

<sup>1</sup>Behavioral and Clinical Neuroscience Institute, Department of Psychology, University of Cambridge, Cambridge, UK; <sup>2</sup>Department of Social and Organizational Psychology and London Institute for Brain and Cognition, London University, London, The Netherlands; <sup>3</sup>Center for Gambling Research at UBC, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada

#### Abstract

This study investigated responses to near-wins (i.e., nonwin outcomes that were close to a major win, and their counterpart, near-losses (nonwin outcomes that are proximal to a major loss)) in a decision-making task, considering (a) task ratings, (b) adjustment of bet amounts, and (c) facial muscle reactivity of zygomaticus and corrugator sites. Compared to full-wins, near-wins decreased self-perceived luck and near-losses increased self-perceived luck, consistent with the effects of spread versus dovetail counterfactual thinking, respectively. Wins and losses both increased zygomaticus reactivity, and losses selectively enhanced corrugator reactivity. Near-wins highlighted zygomaticus activity, but did not affect corrugator activity, thus showing a similar response pattern to actual wins. There was no significant facial EMG effect of near-losses. We infer that near-wins engage some appetitive processing, despite their objective nonwin status.

**Descriptors:** Electromyography, Risk taking, Cognitive distortion, Near-miss, Gambling

Gambling is a widespread form of entertainment where a monetary wager is placed upon the uncertain prospect of a larger monetary win. Its allure can provide insight into the psychological mechanisms of human decision-making. Previous research has shown that near-wins—nonwin outcomes that are proximal to a jackpot— Foster positive feelings (Clay, Cairns, Albert, Devoe, & Ludewig, 2003; Kassirer & Schar, 2001) and increase motivational ratings (Clark, Lawrence, Ashley-Jones, & Gray, 2009). Slot machine near-wins were perceived as being “closer” to wins than to losses (Diamond et al., 2014). Using functional magnetic resonance imaging, near-wins were also found to increase neural signal in brain reward circuitry that overlapped with the jackpot wins (Chase & Clark, 2010; Clark et al., 2009). Nevertheless, near-wins also have a negative emotional component; for example, they are rated as significantly less pleasant than full-miss outcomes (Clark, 2010; Clark et al., 2010, 2012; Qi, Ding, Song, & Yang, 2011).

This work was completed within the University of Cambridge Behavioral and Clinical Neuroscience Institute directorate. TW Robbins, supported by a Clarendon Scholarship Council/Canadian International Scholarship and the Tobacco Damages and Biological Sciences from Dorrance Colquhoun, The Centre for Gambling Research at UBC is supported by an award from the British Columbia Lottery Corporation and the University of British Columbia. We thank Dr. Alexander Kogan for his advice on statistics.

Address correspondence to Yin Wu, Department of Psychology, University of Cambridge, 18B, 18E, 8, Emly Way, 90090-0001.  
\*This is an open access article under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.  
The copyright for this article was changed on 29 September 2014 after original online publication. 159

#### Article

Journal of Conflict Resolution  
00(0), 1–19  
© The Author(s) 2013  
Reprints and permission:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0022002713499718  
jcr.sagepub.com  
SAGE

## Questioning the Effect of Nuclear Weapons on Conflict

Mark S. Bell<sup>1</sup> and Nicholas L. Miller<sup>1</sup>

#### Abstract

We examine the effect of nuclear weapons on interstate conflict. Using more appropriate methodologies than have previously been used, we find that dyads in which both states possess nuclear weapons are not significantly less likely to fight wars, nor are they significantly more or less belligerent at low levels of conflict. This stands in contrast to previous work, which suggests nuclear dyads are some 2.7 million times less likely to fight wars. We additionally find that dyads in which one state possesses nuclear weapons are more prone to low-level conflict (but not more prone to war). This appears to be because nuclear-armed states expand their interests after nuclear acquisition rather than because nuclear weapons provide a shield behind which states can aggress against more powerful conventional-armed states. This calls into question conventional wisdom on the impact of nuclear weapons and has policy implications for the impact of nuclear proliferation.



### Working Paper No. 20 – 2014:

## CAN INFLATION EXPECTATIONS BE MEASURED USING COMMODITY FUTURES PRICES?

Rasheed Saleuddin (corresponding author: [rksms@cam.ac.uk](mailto:rksms@cam.ac.uk)) and D'Maris Coffman

Centre for Financial History, University of Cambridge

### Irregularities in LaCour (2014)

David Broockman, Assistant Professor, Stanford GSB (as of July 1), [dbroockman@stanford.edu](mailto:dbroockman@stanford.edu)  
Joshua Kalla, Graduate Student, UC Berkeley, [kalla@berkeley.edu](mailto:kalla@berkeley.edu)  
Peter Aronow, Assistant Professor, Yale University, [peter.aronow@yale.edu](mailto:peter.aronow@yale.edu)

May 19, 2015

#### Summary

We report a number of irregularities in the replication dataset posted for LaCour and Green (*Science*, “When Contact Changes Minds: An Experiment on Transmission of Gay Equality,” 2014) that jointly suggest the dataset (LaCour 2014) was not collected as described. These irregularities include baseline outcome data that is statistically indistinguishable from a national survey and over-time changes that are unusually small and indistinguishable from perfectly normally distributed noise. Other elements of the dataset are inconsistent with patterns typical in randomized experiments and survey responses and/or inconsistent with the claimed design of the study. A straightforward procedure may generate these anomalies nearly exactly for both studies reported in the paper, a random sample of the 2012 Cooperative Campaign Analysis Project (CCAP) form the baseline data and normally distributed noise are added to simulate follow-up waves.

#### Timeline of Disclosure

- January – April, 2015. Broockman and Kalla were impressed by LaCour and Green (2014) and wanted to extend the article’s methodological and substantive discoveries. We began to plan an extension. We sought to form our priors about several design parameters based on the patterns in the original data on which the paper was based, LaCour (2014). As we examined the study’s data in planning our own studies, two features surprised us: voters’ survey responses exhibit much higher test-retest reliabilities than we have observed in any other panel survey data, and the response and reinterview rates of the panel survey were significantly higher than we expected. We set aside our doubts about the study and invited the authors of our pilot extension to see our pilot extension to see the same parameters. LaCour and Green were both responsive to requests for advice about design details when queried.

Journal of Experimental Political Science 1 (2014) 159–169  
doi:10.1017/xps.2014.9

### Information Spillovers: Another Look at Experimental Estimates of Legislator Responsiveness

Alexander Coppock<sup>1</sup>

#### Abstract

A field experiment carried out by Butler and Nickerson (Butler, D. M., & Nickerson, D. W. (2011). Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Quarterly Journal of Political Science*, 6, 35–83) shows that New Mexico legislators changed their voting decisions upon receiving reports of their constituents’ preferences. The analysis of the experiment did not account for the possibility that legislators may share information, potentially resulting in spillover effects. Working within the analytic framework proposed by Howes et al. (2013), I find evidence of spillovers, and present estimates of direct and indirect treatment effects. The total causal effect of the experimental intervention appears to be twice as large as reported originally.

**Keywords:** Field experiment, spillovers.

#### INTRODUCTION

Butler and Nickerson (2011) report the results of an innovative field experiment testing the responsiveness of legislators to public opinion in New Mexico. Most previous studies of responsiveness note a positive correlation between public opinion and legislators’ choices, which may be due to electoral concerns, the similarity of preferences, or public responsiveness to elite opinion, among many other possible explanations. Butler and Nickerson isolate a single causal channel—the effect of *learning* public opinion on legislators’ voting decisions—by randomly providing some legislators with survey measures of their constituents’ preferences. The headline finding from their study is that representatives change their voting behavior upon acquiring novel public opinion information.

The estimates of responsiveness recovered by Butler and Nickerson (2011) rely on an assumption of non-interference (Cox 1958; Rubin 1980). Legislators respond

The author is grateful to Donald P. Green, Robert Erikson, Gregory Wawro, Peter Aronow, Lindsay Dolan, Albert Fang, and two anonymous reviewers for helpful comments and suggestions, and to Daniel Butler and David Nickerson for providing replication materials.  
<sup>1</sup>Columbia University, New York, NY, USA; e-mail: [ac3242@columbia.edu](mailto:ac3242@columbia.edu)

portion rate of the pilot we could harness the we attempt to contact the staffer at the firm we firm claimed they with the name of the in many aspects of the is 3, 4, 5, and 6 below rd avenues of further ions, Broockman and statistical expertise in (2014). tion and Kalla finds articles 1 and 8 emerge.



# Bringing the Gold Standard into the Classroom: Replication in University Teaching<sup>1</sup>

NICOLE JANZ

*University of Cambridge*

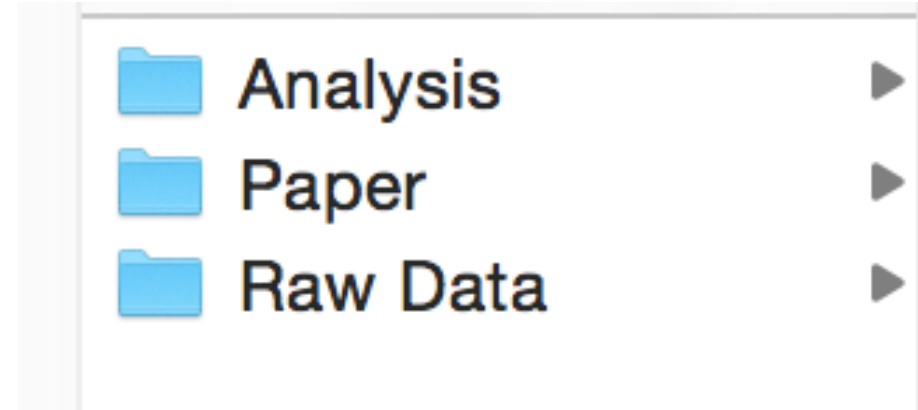
Reproducibility is held to be the gold standard for scientific research. The credibility of published work depends on being able to replicate the results. However, there are few incentives to conduct replication studies in political science. Replications are difficult to conduct, time-consuming, and hard to publish because of a presumed lack of originality. This article sees a solution in a profound change in graduate teaching. Universities should introduce replications as class assignments in methods training or invest in new stand-alone replication workshops to establish a culture of replication and reproducibility. This article will

How to work  
transparently

# Working reproducibly

## Starting out...

- Plan file structure
- Never touch raw data!



## Analysis

- Comment your code
- Keep a log of decisions

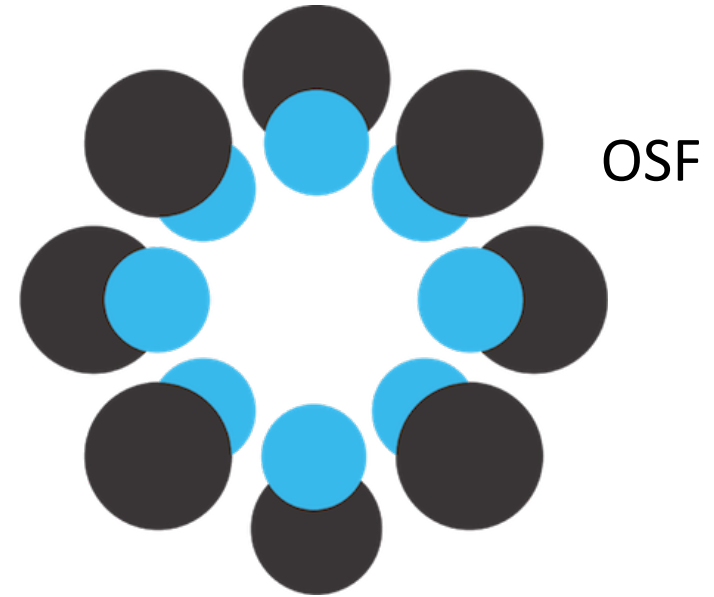
## Writing up

- clarity in methodology section; appendix

**Before you submit:** Replicate your results !!!

# What to share - quantitative

1. Readme file
2. Dataset
3. Software commands
4. Information to reconstruct data



GitHub







# AMERICAN JOURNAL of POLITICAL SCIENCE

*AJPS, South Kedzie Hall, 368 Farm Lane, East Lansing, MI 48824  
ajps@msu.edu, (517) 884-7836*

## GUIDELINES FOR PREPARING REPLICATION FILES

**Version 2.1, May 19, 2016**

**William G. Jacoby  
Robert N. Lupton**

**Michigan State University**

The *American Journal of Political Science* requires the authors of all accepted manuscripts to provide replication files before the article enters the production stage of the publication process. The replication files for each article must be made available as a Dataset (i.e., a collection of files) located in the [AJPS Dataverse](#) on the [Harvard Dataverse Network](#). Instructions for getting started on the *AJPS* Dataverse can be found in the [“Quick Reference for Uploading Replication Files,”](#)

# Support your claims - qualitative

satellite images, interview transcripts, personal diaries, video clips, newspaper articles, speeches...

## Transparency Appendix:

- how you **evaluated** persuasiveness & consistency of evidence
- logic and **steps** in process tracing
- Upload **files** and **fragments** e.g. partial transcripts (100-150 w.)



# When to protect the data

- Confidential / proprietary data
- Protect individuals
- Informed consent obtained?



- **Anonymization**
- **Justify** why you withhold data



What's in it  
**for me?**



avoid  
**disaster**



easier to  
**write up**



persuade  
**reviewers**



enables  
**continuity**



build your  
**reputation**



# Get in Touch



Twitter: @PolSciReplicate



<http://PoliticalScienceReplication.wordpress.com/>



Vising scholar 2016  
and Catalyst



Ambassador

# Useful resources

# Materials - Replication

- King, Gary. (2006). **How to Write a Publishable Paper as a Class Project**, copy at: <http://gking.harvard.edu/papers>
- Janz, N. (2015) **Bringing the Gold Standard Into the Class Room: Replication in University Teaching**, International Studies Perspectives, Article first published online: 9 March 2015. Copy at: <http://tinyurl.com/q2qnrvn>
- Brandt et al. (2014) **The Replication Recipe: What makes for a convincing replication?** Journal of Experimental Social Psychology, Vol 50, pp. 217-224. Copy at: <http://tinyurl.com/poe474k>

# Materials – Transparent Workflow

- Christensen, Garret (2016). **Manual of Best Practices in Transparent Social Science Research**  
<https://github.com/garretchristensen/BestPracticesManual>
- Open Science Framework. **Transparency and Openness Promotion (TOP) Guidelines**. <https://cos.io/top/>
- **TIER Documentation Protocol**  
<https://www.haverford.edu/project-tier/protocol-v2>

# Adding value to a replication

1. Theoretical contribution: questioning the arguments
2. Statistical contribution

## **Sample size:**

Power calculations (how big should the sample be?)

More years, more countries (units)

New samples (experiments)

Different subsets of your data set (e.g. only OECD countries)

Missing data handling (multiple imputation)

## **Changing measurements:**

Change of variables: %GDP, log transformation, different ways of dealing with negative values for logging, different measurement for the same variable

## **Model specification:**

Standard errors treatment, LDV, lags

Interactions

Dummy variables

Omitted variables

Reversed causality

Adjusted / improved / advanced models

## **Robustness/Sensitivity checks:**

How much do betas and standard errors change when we change model specifications? Are they very 'sensitive' even to small changes/outliers?