Solving the Reproducibility Crisis

A teaching perspective

Dr Nicole Janz
Assistant Professor in International Relations

March 1, 2018

Cambridge Statistics Discussion Group

Most scientists 'can't replicate studies by their peers'

By Tom Feilden
Science correspondent, Today programme

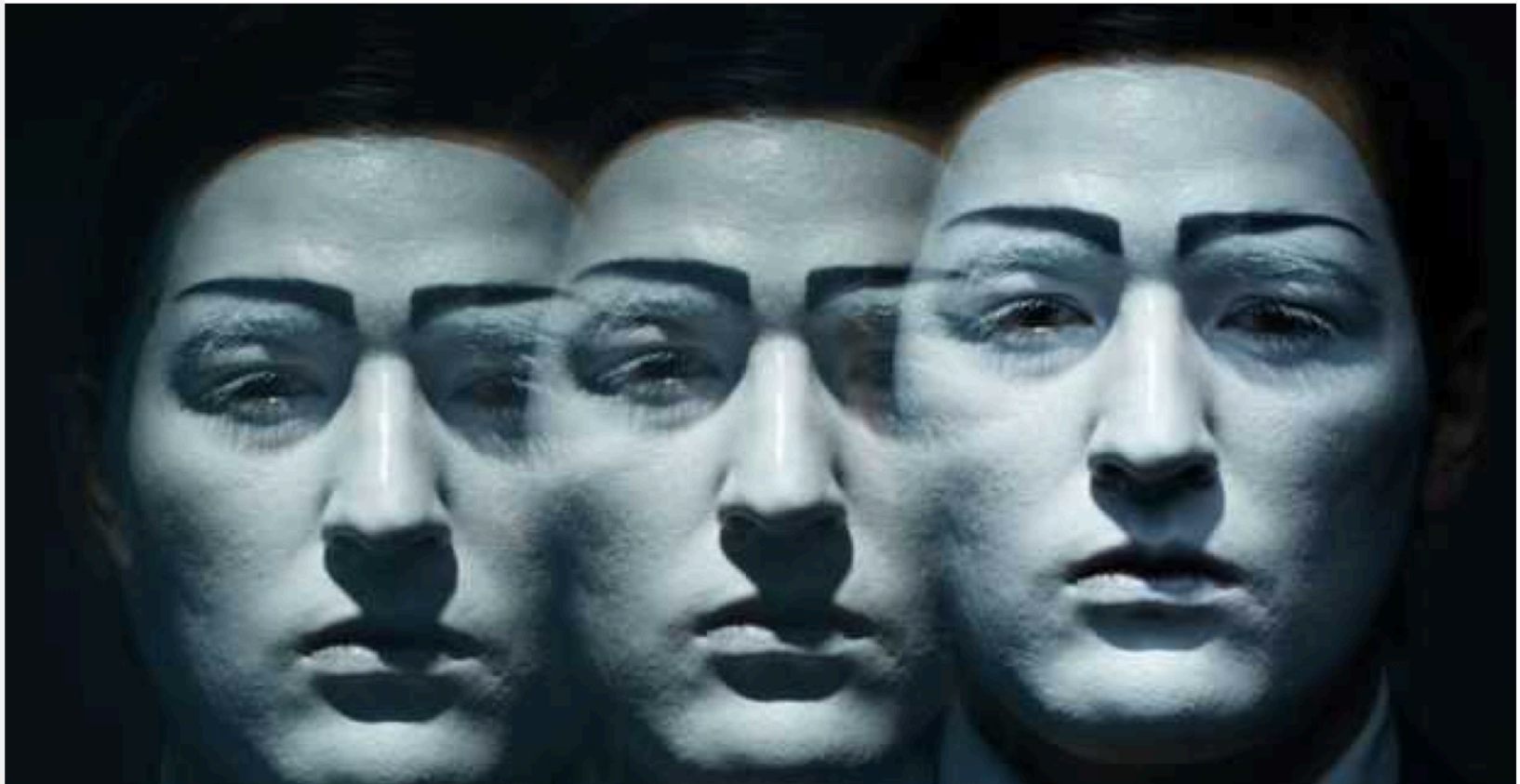
22 February 2017 | Science & Environment

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



1

Many studies are **not reproducible**

2

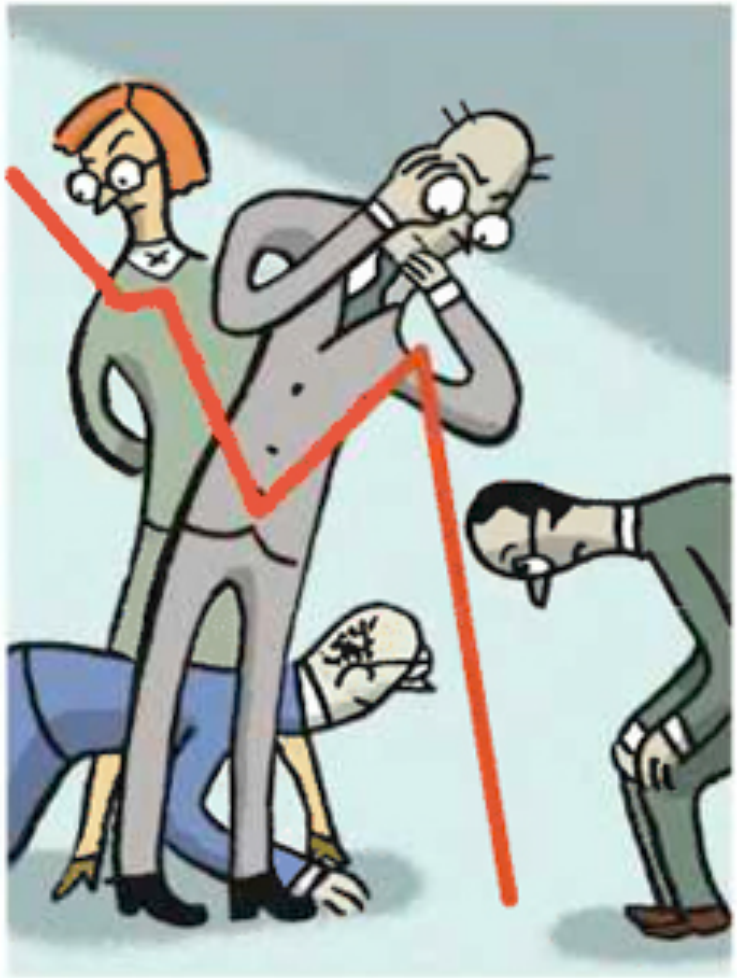
Authors

don’t share

data

THE REINHART AND ROGOFF CONTROVERSY: A SUMMING UP

By John Cassidy April 26, 2013



In one of life's little ironies, last Friday's disappointing G.D.P. figures, which reflected a sharp fall in government spending, appeared on the same day that the economists Carmen Reinhart and Kenneth Rogoff published an Op-Ed in the *Times* defending their famous (now infamous) research that conservative politicians around the world had seized upon to justify penny-

May 29, 2015 12:34 p.m.

The Case of the Amazing Gay-Marriage Data: How a Graduate Student Reluctantly Uncovered a Huge Scientific Fraud

By **Jesse Singal**

NEW YORK

3

Few journals **enforce data upload**

RESEARCH

data availability in political science journals

sergiu gherghina[a,*] *and alexia katsanidou*[a]

[a]GESIS Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Cologne, Germany.

E-mails: sergiu.gherghina@gesis.org; alexia.katsanidou@gesis.org

*Corresponding author.

advance online publication, 1 March 2013; doi:10.1057/eps.2013.8

Abstract

A characteristic of recent decades of scholarly work in the social sciences has been the increased amounts of empirical research. Access and availability of data are prerequisites for further research, replication work, and scientific development. As international peer-reviewed journals have gradually become the central forum for research debate, moves towards data sharing are dependent upon the policies of journals regarding data availability. This article examines contemporary data availability policies in political science and investigates the extent to which journals adopt such policies and their content. It also identifies a few factors associated with the existence of such policies.

FEBRUARY 16TH 2017

Replication in international relations

BY NILS PETTER GLEDITSCH AND NICOLE JANZ

The integrity of science is threatened in many ways – by direct censorship; by commercial, political, or military secrecy; by various forms of publication bias; by exorbitant journal subscription fees that effectively deny access to the general public; by cheating and falsification of results; and by sloppiness in the research process or the editorial process prior to publication. There isn't a single antidote to all these problems, but transparency goes a long way in relation to many of them.

Most scientific knowledge remains unchecked.

Solutions

1. Journals

SCIENTIFIC STANDARDS

Promoting an open research culture

Author guidelines for journals could help to promote transparency, openness, and reproducibility

By **B. A. Nosek,* G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. A. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni**

Transparency, openness, and reproducibility are readily recognized as vital features of science (*1*, *2*). When asked, most scientists embrace these features as disciplinary norms and values (*3*). Therefore, one might expect that these valued features would be routine in daily practice. Yet, a growing body of evidence suggests that this is not

Summary of the eight standards and three levels of the TOP guidelines

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Preregistration of studies	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.

2. Funders

Proposals **must** include **Data Management Plan.**

- Describe **data** and **access**
- How data will be **archived** for re-use

https://www.nsf.gov/eng/general/dmp.jsp

Publicly funded **research data are a public good**

- Make **data discoverable** and enable **effective re-use**

http://www.rcuk.ac.uk/research/datapolicy/

3. Promoting benefits

COMMENT Open Access

Five selfish reasons to work reproducibly

Florian Markowetz

Abstract

And so, my fellow scientists: ask not what you can do for reproducibility; ask what reproducibility can do for you! Here, I present five reasons why working reproducibly pays off in the long run and is in the self-interest of every ambitious, career-oriented scientist.

Keywords: Reproducibility, Scientific career

A complex equation on the left half of a black board, an even more complex equation on the right half. A short sentence links the two equations: “Here a miracle oc-

how science actually is. And, whether you like it or not, science is all about more publications, more impact factor, more money and more career. More, more, more… so how does working reproducibly help me achieve more as a scientist.

Reproducibility: what’s in it for me?

In this article, I present five reasons why working reproducibly pays off in the long run and is in the self-interest of every ambitious, career-oriented scientist.

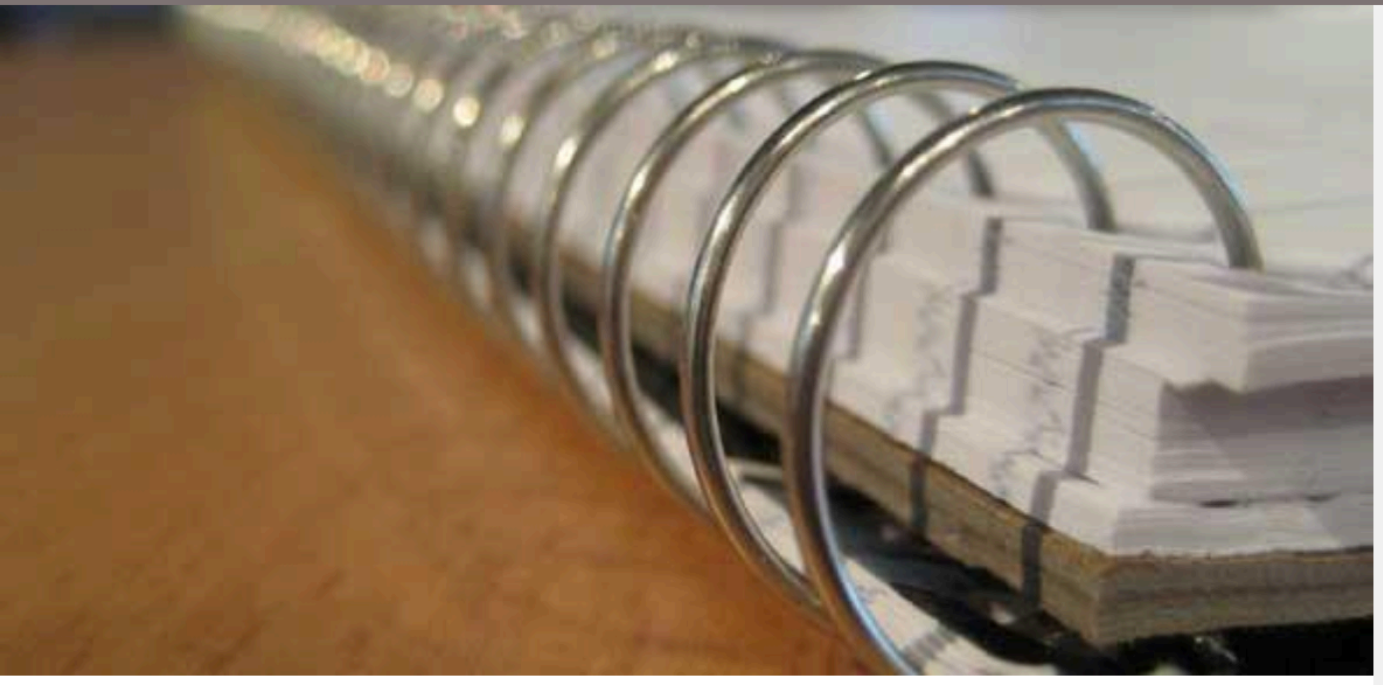
Reason number 1: reproducibility helps to avoid disaster

5 selfish reasons
to work reproducibly

1. Avoid disaster
2. Easier to write papers
3. Easier to talk to reviewers
4. Continuity of your work/in the lab
5. Reputation

4. University teaching

Bringing the gold standard into the classroom: teaching replication and reproducibility at Cambridge



Published

13 Feb 2014

Image

Notebook rings

Credit: Brenderous

Share

Email 1 reddit 0

Share 41 Share 1

Tweet 47 Share 480

Like 26

Quality standards in the sciences have recently been heavily criticised in the academic community and the mass media. Scandals involving fraud, errors or misconduct have stirred a debate on reproducibility that calls for fundamental changes in the way research is done. As a new teaching course at Cambridge shows, the best way to bring about change is to start in the classroom, explains course instructor Nicole Janz.

Why should you replicate?

Learn Statistics

- Real life data
- Author decisions
- Bugs included
- More fun than textbook

Reproducibility routine

- When are published results really reproducible?

Publish

- Add value
- Publish faster

How to replicate a study

Four main challenges in replication research

Challenge 1 Too many **definitions** of replication

Challenge 2 Too many articles – **which one** to pick?

Challenge 3 How to replicate **systematically**

Challenge 4 **Publishing** a replication study

Challenge 1: Too many definitions of replication

Use terminology accepted in **your** field

Political Science (see King 2003)

Duplication	Replication
Verify research results	Test the robustness of the original research results
exact same data set	***new data***
exact same methods	***new models***

In Psychology this would be…

Close replication	Conceptual replication
Verify research results by following original study **as exactly as possible** (participant recruitment, measurements, procedures, and analyses). Ideally the **only differences** between the two are the inevitable ones (e.g. different participants).	Test the robustness of the original study **providing new tests of a theory** (Simons 2014) Conceptual replications **assume the validity** of the original finding and its explanation and **test a generalization** of it (Larzelere et al. 2014)

JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12139

THE MEANING OF FAILED REPLICATIONS: A REVIEW AND PROPOSAL

Michael A. Clemens*

Center for Global Development and IZA

Abstract. The welcome rise of replication tests in economics has not been accompanied by a consensus standard for determining what constitutes a *replication*. A discrepant replication, in current usage of the term, can signal anything from an unremarkable disagreement over methods to scientific incompetence or misconduct. This paper proposes a standard for classifying one study as a replication of some other study. It is a standard that places the burden of proof on a study to demonstrate that it should have obtained identical results to the original, a conservative standard that is already used implicitly by many researchers. It contrasts this standard with decades of unsuccessful attempts to harmonize terminology, and argues that many prominent results described as replication tests should not be described as such. Adopting a conservative standard like this one can improve incentives for researchers, encouraging more and better replication tests.

Keywords. Ethics; Open data; Replication; Robustness; Transparency

Table 2. Correspondence between Table 1 and Current Terminology.

	Replication test*	Robustness test*	Source
Economics	Type I **replication**	Type II, III, IV **replication**	Mittelstaedt and Zorn, 1984
	Econometric audit	Improvisational **replication**	Kane, 1984
	"Reproduction" **replication**	"Reexamination" **replication**	Fuess, 1996
	"Reproduction" **replication**	"Robustness" **replication**	Kniesner, 1997
	Replication of the first degree	Higher order **replication**/reanalysis	Arulampalam *et al.*, 1997
	"Narrow sense" **replication**	"Wide sense" **replication**	Pesaran, 2003
	Pure **replication**	Statistical/Scientific **replication**	Hamermesh, 2007
	Replication	—	McCullough *et al.*, 2008
	Replication	Stress test	Vinod, 2009
	Replication/reproduction	—	Koenker and Zeileis, 2009
	Repeatability/Strict **replication**	Conceptual **replication**	Ioannidis and Doucouliagos, 2013
	Replication	—	Data policy of *AER*, *JPE*, etc.
Statistics	Close **replication**	Differentiated **replication**	Lindsay and Ehrenberg, 1993
	Computational reproduction	—	Donoho, 2010
	Replication/reproduction	—	Stodden, 2010
	Reproduction	**Replication**	Peng, 2011
Political science	**Replication**	Extension	King, 1995
	"Verification" reanalysis	"**Replication**" reanalysis	Herrnson, 1995
	Replication	Extension, improvement	King, 2006
	Narrow **replication**	Broad **replication**	Dafoe, 2014
Sociology	Retest/internal **replication**	Independent/theoretical **replication**	La Sorte, 1972
	Identical **replication**	Virtual/systematic **replication**	Finifter, 1972
	Replication type *a*	**Replication** type *b* . . . *p*	Bahr *et al.*, 1983
	Repetition/checking	**Replication**	Collins, 1991
	Replication	Reproduction, robustness	Cartwright, 1991
Psychology	Literal/operational **replication**	Constructive **replication**	Lykken, 1968
	Replication	Quasi-**replication**	Cronbach, 1975
	Exact **replication**	Partial/conceptual **replication**	Hendrick, 1990
	Internal **replication**	External **replication**	Thompson, 1994
	Direct **replication**	Conceptual **replication**	Schmidt, 2009
	Exact **replication**	Close **replication**	Brandt *et al.*, 2014
Business	Experimental **replication**	Nonexperimental/corroboration **replication**	Leone and Schultz, 1980
	Perfect **replication**	Imperfect **replication**	Farley *et al.*, 1981
	Replication	Extension	Hubbard and Armstrong, 1994
	Strict **replication**	Significant sameness	Barwise, 1995
	Duplication	Operational **replication**	Madden *et al.*, 1995
	Checking	**Replication**, reanalysis, extension, etc.	Tsang and Kwan, 1999
	Strict **replication**	Partial/conceptual **replication**	Darley, 2000
	Replication	Extension	Easley and Madden, 2000
	Type 0, I **replication**	Type II, III **replication**	Easley *et al.*, 2000
	Statistical **replication**	Scientific **replication**	Hunter, 2001
	Replication	**Replication** with extension	Evanschitzky *et al.*, 2007

Four main challenges in replication research

Challenge 1 Too many **definitions** of replication

Challenge 2 Too many articles – **which one** to pick?

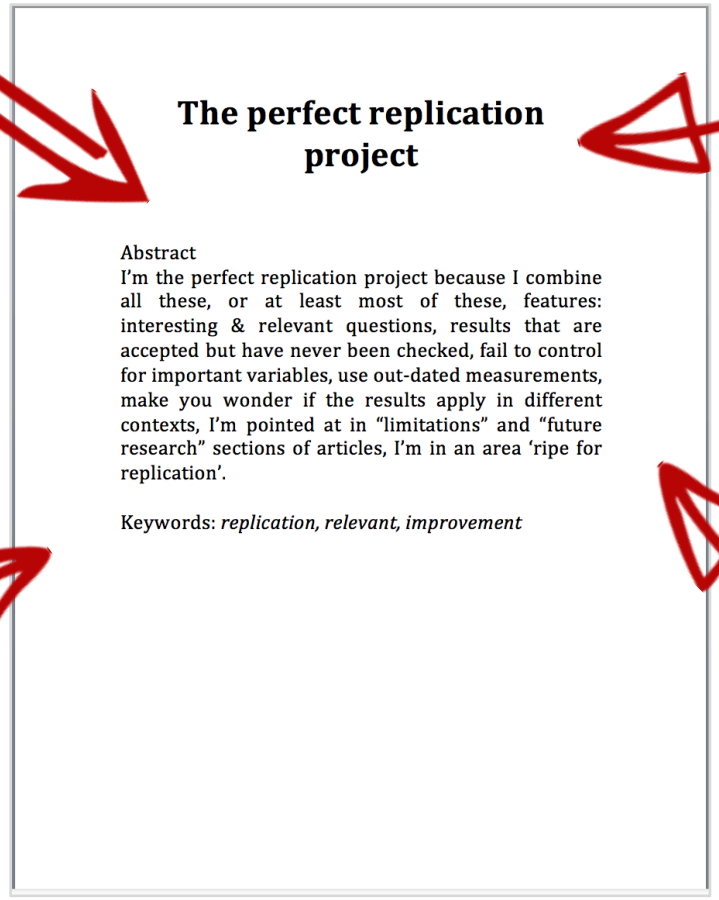
Challenge 3 How to replicate **systematically**

Challenge 4 **Publishing** a replication study

Challenge 2: 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.[1] 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.[2] 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: creinhar@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.

[1] 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.

[2] 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.

This content downloaded from 131.111.194.10 on Wed, 03 Jun 2015 14:40:40 UTC
All use subject to JSTOR Terms and Conditions

Practical tips for choosing a study

"By far the biggest problem was that some students picked papers that were too difficult given their background."

Matthew Salganik, Department of Sociology, Princeton University

- Don't select a study with **methods** that you don't know or can't learn quickly
- Study not older than **5 years** and from a **good journal**
- **Data** (and ideally software **code**) should be available

Four main challenges in replication research

Challenge 1 Too many **definitions** of replication

Challenge 2 Too many articles – **which one** to pick?

Challenge 3 How to replicate **systematically**

Challenge 4 **Publishing** a replication study

Challenge 3: How to replicate systematically

Project Plan

1. Summary of main **results** to be replicated
2. List of main statistical **methods** you'd have to learn
3. Summary of **data** availability and access
4. Is the software **code** online?
5. List of ideas for how the paper could be **extended** with new data and methods

(from *Matthew Salganik, Princeton*)

Tip from my students: word document with

- screenshots of all **tables** and **figures** in original paper
- Copy paste models and results **description**

Practical steps in a replication study

1 Select paper

2 Access data & code

3 Identify each variable

2-3 weeks

4 Reproduce tables, figures

5 Compare

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!**

Adding value to a duplication

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)

Model specification:
Standard errors treatment, LDV, lags
Interactions
Dummy variables
Omitted variables
Reversed causality
Adjusted / improved / advanced models

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

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?

Comparing your results with the original study

Clarify with precision the extent to which you were able to replicate the author's results.
Gary King (2006)

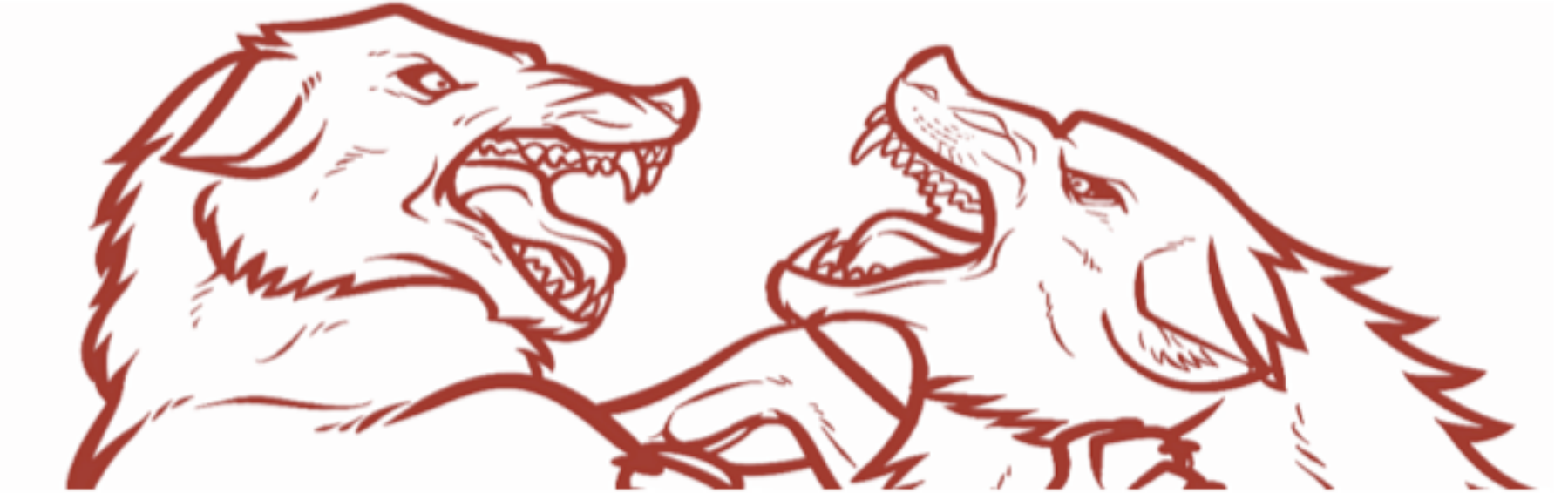
- A replication can fail at **different stages**.
- Exact same data and methods: results cannot be **duplicated**.
- New data, models, methods: you have to describe **exactly** in which **way**, and at which **step**, the replication has failed
- Different measurements of concepts that are hard to operationalize, e.g. human rights, can **naturally yield different results**!

What exactly failed to replicate?

Checklist:

- Could you not identify which **variable** is which in the original data?
- Was a **transformation** of variables in the original data set unclear?
- Were there errors in the original **data set**?
- How did the coefficients, standard errors, confidence intervals differ?
- How did the **figures** look different after replicating them?
- Did a small change in **outlier** treatment change the results?
- How did you measure the variables **differently** when ‘adding value’?
- Did you **update** the data (e.g. for the recent years or more countries) and the results changed?

Communicating failed replications



Be professional!

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).

Four main challenges in replication research

Challenge 1 Too many **definitions** of replication

Challenge 2 Too many articles – **which one** to pick?

Challenge 3 How to replicate **systematically**

Challenge 4 Publishing a replication study

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

Quarterly Journal of Political Science, 2010, 5: 339–356

Voting Costs and Voter Turnout in Competitive Elections

Bernard Fraga[1] and Eitan Hersh[2,*]

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

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[1] and Nicholas L. Miller[1]

… we analyze a dyad-year data set (used by Rauchhaus 2009) to examine whether existing findings on the effect of symmetric nuclear weapons possession on conflict are robust to the improvements noted above. We find that once pre-nuclear dyadic conflict is controlled for, symmetric nuclear dyads are not more likely to experience low-level conflict.

Political Regimes and International Trade: The Democratic Difference Revisited

XINYUAN DAI *University of Illinois at Urbana–Champaign*

How 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 hi 0). Cont r a demo ed. Thus, er level

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

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.

How tra de R model. than grou offers to both demo de alter- questions, alculate analyze a no clear- in which ring the ers agreed autocratic

To distin between field, Milne country argue that A, or a approval o gislature while an au try. The TILI games played by democratic pairs, by autocratic ideal level of trade barriers at home and abroad for

Public Finance Review
2015, Vol 43(2) 155-178
© The Author(s) 2014
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1091142113510197
pfr.sagepub.com

A Replication of "Economic Development and the Impacts of Natural Disasters" (*Economics Letters* 2007)

Robert Mercer[1] and W. Robert Reed[2]

Abstract

This study replicates the empirical findings of Toya and Skidmore (2007), henceforth TS, and performs a variety of robustness checks. Using an extensive data set of international disasters, TS report that a number of

5/

Journals Open to Replication (selection)

Political Science

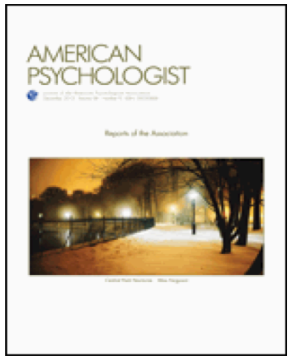
Research & Politics ^

*

*

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

The ReScience Journal

ABOUT READ WRITE EDIT BOARD FAQ

Reproducible Science is good. Replicated Science is better.

ReScience is a peer-reviewed journal that targets computational research and encourages the explicit replication of already published research, promoting new and open-source implementations in order to ensure that the original research is reproducible.

To achieve this goal, the whole publishing chain is radically different from other traditional scientific journals. ReScience lives on GitHub where each new implementation of a computational study is made available together with comments, explanations and tests. Each submission takes the form of a pull request that is publicly reviewed and tested in order to guarantee that any researcher can re-use it. If you ever replicated computational results from the literature in your research, ReScience is the perfect place to publish your new implementation.

ReScience is collaborative by design. Everything can be forked and modified. Don't hesitate to write a submission, join us and to become a reviewer.

http://rescience.github.io/

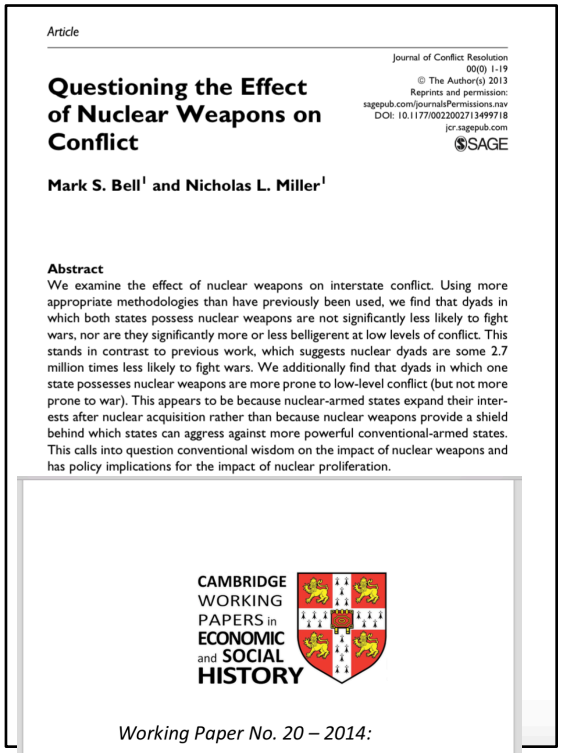
Replications by Early Career Researchers


POLITICAL ECONOMY RESEARCH INSTITUTE
PERI
University of Massachusetts Amherst

POLITICAL ECONOM
RESEARCH INSTITUT

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

Thomas Herndon, Michael Ash and Robert Pollin


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[1] and Nicholas L. Miller[1]

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.


Irregularities in LaCour (2014)

David Broockman, Assistant Professor, Stanford GSB (as of July 1), dbroockman@stanford.edu
Joshua Kalla, Graduate Student, UC Berkeley, kalla@berkeley.edu
Peter Aronow, Assistant Professor, Yale University, 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 support for 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 awaited the launch of our pilot extension to see if we could manage the same parameters. LaCour and Green were both responsive to requests for advice about design details when queried.


Psychophysiology, 52 (2015), 359–366. Wiley Periodicals, Inc. Printed in the USA.
Copyright © 2014 The Authors. Psychophysiology published by Wiley Periodicals, Inc. on behalf of Society for Psychophysiological Research
DOI: 10.1111/psyp.12336

PSYCHOPHYSIOLOGY

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

YIN WU,[a] ERIC VAN DIJK,[b] AND LUKE CLARK[a,c]
[a]Behavioural and Clinical Neuroscience Institute, Department of Psychology, University of Cambridge, Cambridge, UK
[b]Department of Social and Organizational Psychology and Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands
[c]Centre 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, measuring (a) luck ratings, (b) adjustment of bet amount, and (c) facial muscle reactivity at zygomaticus and corrugator sites. Compared to full-misses, near-wins decreased self-perceived luck and near-losses increased self-perceived luck, consistent with the effects of upward versus downward counterfactual thinking, respectively. Wins and losses both increased zygomaticus reactivity, and losses selectively enhanced corrugator reactivity. Near-wins heightened zygomaticus activity, but did not affect corrugator activity, thus showing a similar response pattern to actual wins. There were no significant facial EMG effects of near-losses. We infer that near-wins engender 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 persistent play (Côté, Caron, Aubert, Desrochers, & Ladouceur, 2003, Kassinove & Schare, 2001) and increase motivational ratings (Clark, Lawrence, Astley-Jones, & Gray, 2009). Slot machine near-wins were perceived as being "closer" to wins than to losses (Dymond 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., 2009, 2013; Qi, Ding, Song, & Yang, 2011).

Psychophysiology may provide a useful tool for further characterizing the bivalent emotional response to these events. Past work has shown that near-wins increase electrodermal activity (EDA) and heart rate acceleration, in comparison to full-miss outcomes (Clark, Crooks, Clarke, Aitken, & Dunn, 2012; Clark et al., 2013; Dixon et al., 2011). EDA is a sensitive marker of physiological arousal that responds to both aversive and appetitive stimuli, and thus offers limited valence specificity (Dawson, Schell, & Filion, 2007; Lobbestael, Arntz, & Wiers, 2008). Phasic heart rate changes do show some differences by valence (e.g., Bradley, Codispoti, Cuthbert, & Lang, 2001), but this is a complex multiphasic response that also varies across individuals (Bodes, Cook, & Lang, 1985). Facial muscle activity offers an alternative probe of stimulus-evoked emotional reactivity with superior valence differentiation, with zygomaticus activity (recorded on the cheek) linked to appetitive processing, and corrugator supercilii activity (recorded on the eyebrow) scaling with aversive processing (Cacioppo, Petty, Losch, & Kim, 1986; Lang, Greenwald, Bradley, & Hamm, 1993). The present study employed facial electromyography (EMG) at these two sites to better decompose the bivalent emotional nature of gambling "near" events.

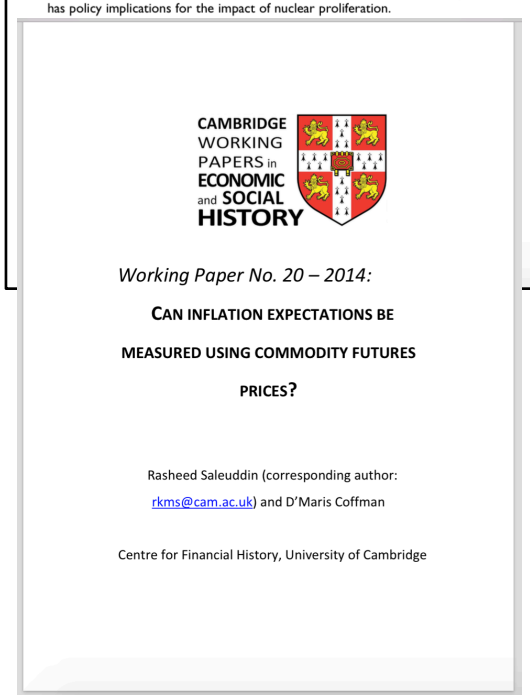
Little attention has been paid to the natural counterpart to the near-win, the "near-loss." In a decision with the possibility of losing money, what happens if you discover that you narrowly missed a major loss? While these events are less ubiquitous in gambling behavior, they do occur across many areas of day-to-day decision making, for example, when we narrowly miss an accident, or traffic jam, and they have received some attention in occupational psychology. For example, narrowly avoiding a great disaster lowered the future perceived risk of that event occurring, and increased risky choice (Dillon & Tinsley, 2008). The present study

This work was completed within the University of Cambridge Behavioural and Clinical Neuroscience Institute (director: TW Robbins), supported by a consortium award from the Medical Research Council (MRC Ref G1000183) and Wellcome Trust (WT Ref 093875/Z/10/Z). YW was supported by a Chinese Scholarship Council-Cambridge International Scholarship and the Treherne Studentship in Biological Sciences from Downing College, Cambridge. The Centre for Gambling Research at UBC is supported by an award from the British Columbia Lottery Corporation and the British Columbia Government. We thank Dr. Aleksandr Kogan for his advice on statistics.

Address correspondence to: Yin Wu, Department of Psychology, University of Cambridge, Cambridge, CB2 3EB, UK. E-mail: yw330@cam.ac.uk

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

[The copyright line for this article was changed on 29 September 2014 after original online publication.]

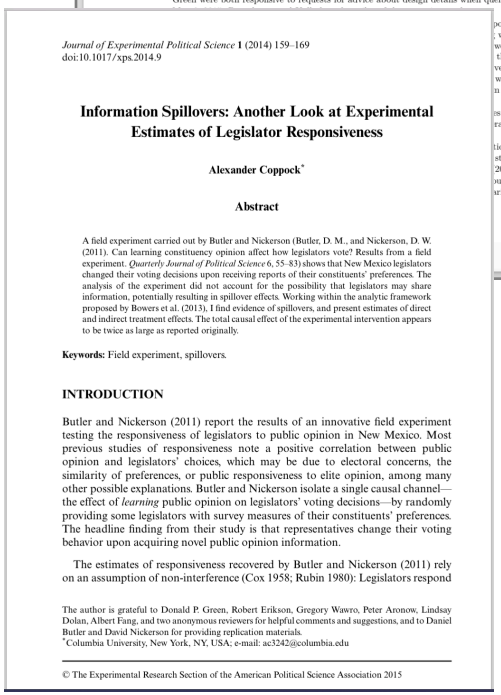

CAMBRIDGE WORKING PAPERS in ECONOMIC and SOCIAL HISTORY

Working Paper No. 20 – 2014:

CAN INFLATION EXPECTATIONS BE MEASURED USING COMMODITY FUTURES PRICES?

Rasheed Saleuddin (corresponding author: rkms@cam.ac.uk) and D'Maris Coffman

Centre for Financial History, University of Cambridge


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*

Abstract

A field experiment carried out by Butler and Nickerson (Butler, D. M., and Nickerson, D. W. (2011). Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Quarterly Journal of Political Science* 6, 55–83) 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 Bowers 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.

*Columbia University, New York, NY, USA; e-mail: ac3242@columbia.edu

© The Experimental Research Section of the American Political Science Association 2015

Curriculum:

- Workshops: reproducible workflow
- Replications as class assignment in stats course
- Replication projects for students

PIs:

- establish a culture of reproducibility & replication within your lab
- lab members replicate each other before journal submission
- cross-check your code

International
Studies
Perspectives

International Studies Perspectives (2015), 1–16.

Bringing the Gold Standard into the Classroom: Replication in University Teaching[1]

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

Photo: Sam Mather

Thank you!

nicole.janz@nottingham.ac.uk

@polscireplicate

Political Science Replication Blog

Materials

- 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

- Markowetz, Florian (2015), **Five selfish reasons to work reproducibly**. Genome Biology 16:274.

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

- Janz, Nicole & Figueiredo, Dalson (2017, March 13). **Workshop: The Gold Standard of Reproducible Research**. Retrieved from https://osf.io/2fqnw/ (slides, handouts)

Replication Recipe by Brandt et al 2014

- A good source, particularly for Psychologists, is this replication recipe:

A 36-question guide to the Replication Recipe.

The Nature of the Effect

1. Verbal description of the effect I am trying to replicate:
2. It is important to replicate this effect because:
3. The effect size of the effect I am trying to replicate is:
4. The confidence interval of the original effect is:
5. The sample size of the original effect is:
6. Where was the original study conducted? (e.g., lab, in the field, online)
7. What country/region was the original study conducted in?
8. What kind of sample did the original study use? (e.g., student, Mturk, representative)
9. Was the original study conducted with paper-and-pencil surveys, on a computer, or something else?

Designing the Replication Study

10. Are the original materials for the study available from the author?
    a. If not, are the original materials for the study available elsewhere (e.g., previously published scales)?
    b. If the original materials are not available from the author or elsewhere, how were the materials created for the replication attempt?
11. I know that assumptions (e.g., about the meaning of the stimuli) in the original study will also hold in my replication because:
12. Location of the experimenter during data collection:
13. Experimenter knowledge of participant experimental condition:
14. Experimenter knowledge of overall hypotheses:
15. My target sample size is:
16. The rationale for my sample size is:

Documenting Differences between the Original and Replication Study

For each part of the study indicate whether the replication study is Exact, Close, or Conceptually Different compared to the original study. Then, justify the rating.

17. The similarities/differences in the instructions are: