

“See For Yourself”

The **Pleasures** and **Sorrows** of Transparency in Social Science Research

Cristobal Young

Department of Sociology, Stanford University



THE ROYAL SOCIETY

(3075)

Num. 20.

PHILOSOPHICAL TRANSACTIONS.

February 19. 1677.

THE CONTENTS.

A Letter of Mr. Isaac Newton, Mathematick Professor in the University of Cambridge, concerning his New Theory about Light and Colors: Where Light is declared to be not Similar or Homogeneous, but consisting of diverse rays, some of which are more refrangible than others: And Colors are affirmed to be not Qualifications of Light, derived from Reflections of natural Bodies, (as the generally believed,) but Original and Connate properties, which in diverse rays are diverse: Where several Observations and Experiments are added to prove the said Theory. An Account of some Books; I. A Description of the EAST-INDIAN COASTS, MALABAR, COROMANDEL, CEYLON, &c. in Dutch, by Phil. Baldaeus. II. Arsoni le Grand INSTITUTIO PHILOSOPHICÆ, secundum principia Renati Des-Cartes, nova methodo abornata & explicata. III. An Essay to the Advancement of MUSICK, by Thomas Simon M. A. Advertisements about THOMAS SYMPSON. An Index for the Trade of the Year 1677.

A Letter of Mr. Isaac Newton, Professor of the Mathematics in the University of Cambridge, containing his New Theory about Light and Colors: sent by the Author to the Publisher from Cambridge, 11th. 6. 1677, in order to be communicated to the R. Society.

SIR,

TO perform my late promise to you, I shall without further ceremony acquaint you, that in the beginning of the Year 1666 (at which time I applied my self to the grinding of Optick glasses of other figures than Spherical) I procured me a Triangular glass-Prism, to try therewith the celebrated Phænomena of

O R S R

Colors.

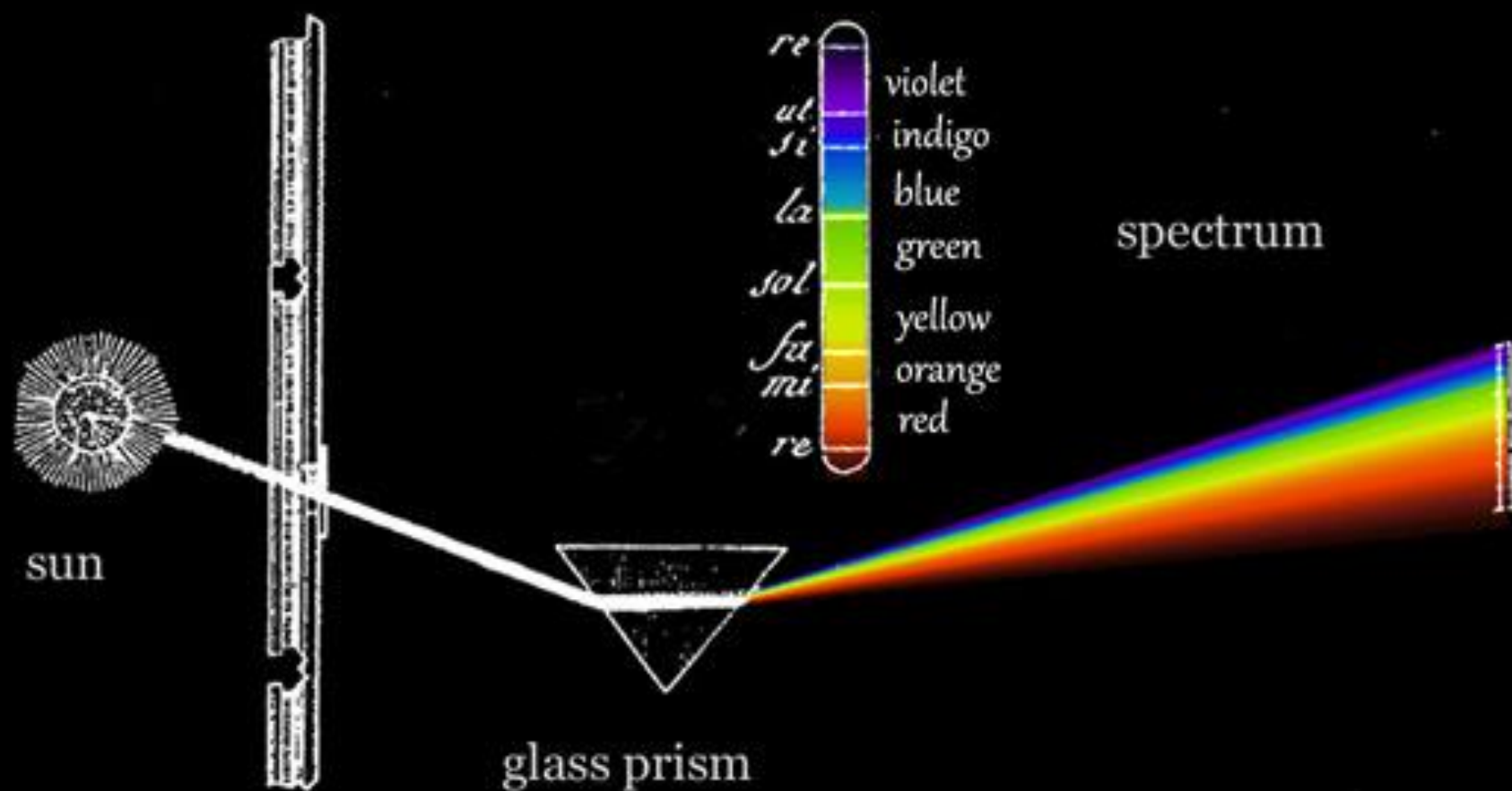


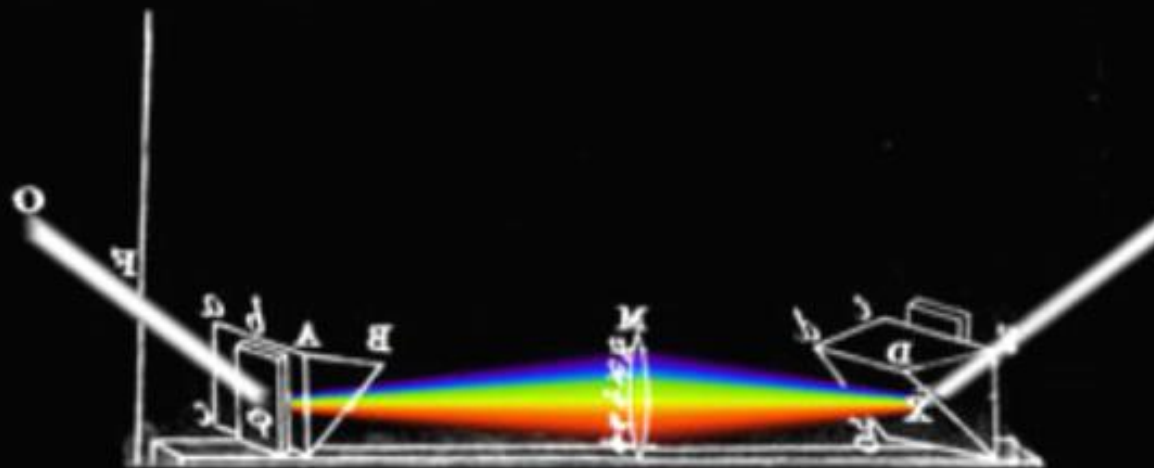
Isaac Newton and the Theory of **Color**

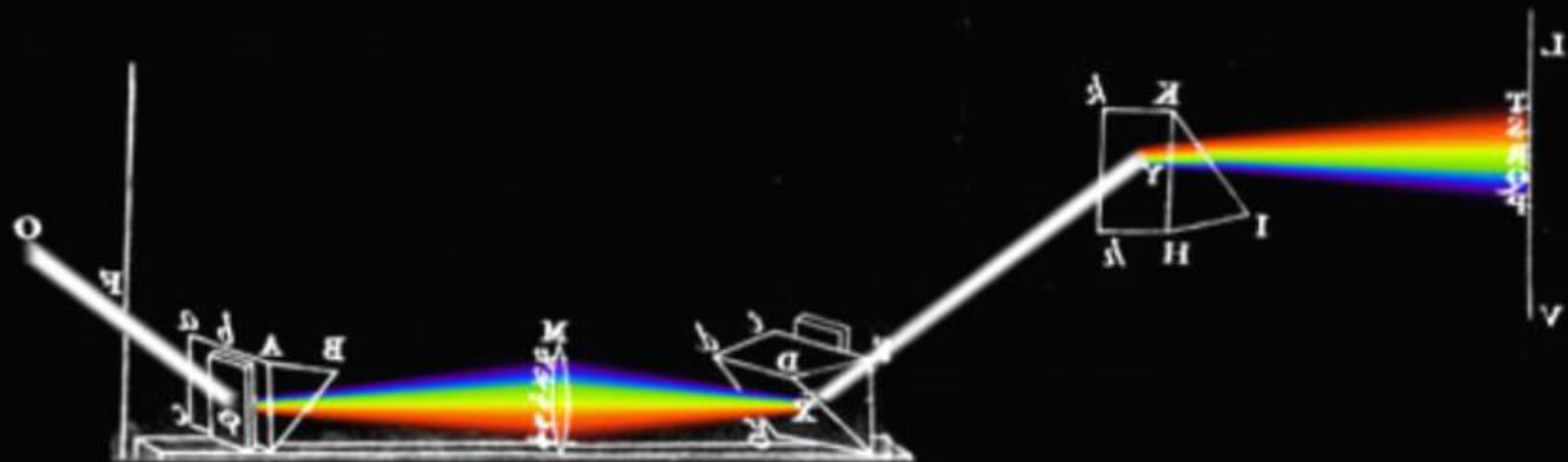
Journal of the Royal Society, 1671



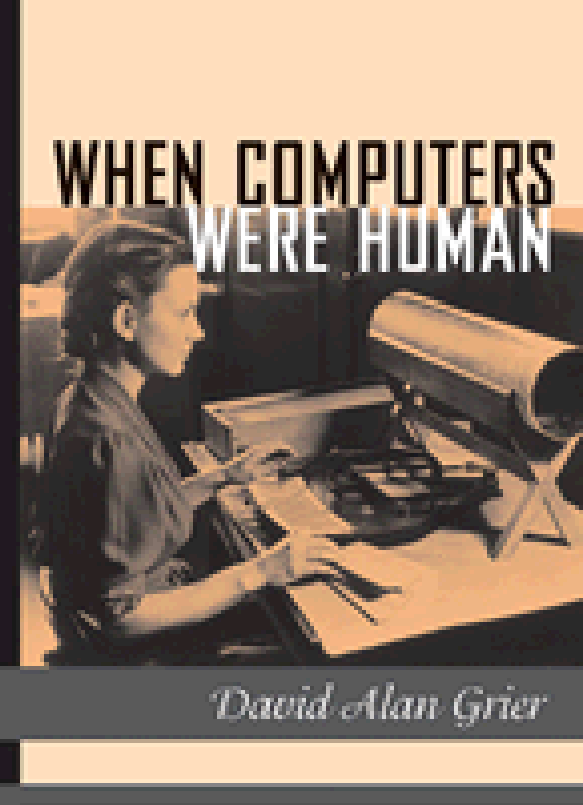








Computer Lab, circa 1930



Running many different models was computationally infeasible

IBM 360

68234ASHLEY COMPANY 2911 S. T

CUST NO	CUSTOMER NAME

A few days per regression command...

68234 HAWLEY COMPANY										2911 S. TREMONT ST.										AUSTIN TX.										092740										98766										82509																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
CUST NO		CUSTOMER NAME		STREET ADDRESS		CITY AND STATE		INVOICE DATE		INVOICE NO		INVOICE AMOUNT		CUST NO		CUSTOMER NAME		STREET ADDRESS		CITY AND STATE		INVOICE DATE		INVOICE NO		INVOICE AMOUNT																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													

Replication in Sociology

Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis

Cristobal Young¹ and Katherine Holsteen²

Sociological Methods & Research

2017, Vol. 46(1) 3-40

© The Author(s) 2015

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0049124115610347

journals.sagepub.com/home/smr



Replication in Sociology

Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis

Cristobal Young¹ and Katherine Holsteen²

Sociological Methods & Research

2017, Vol. 46(1) 3-40

© The Author(s) 2015

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0049124115610347

journals.sagepub.com/home/smr



Replication package in Stata:

do http://web.stanford.edu/~cy10/public/mrobust/install_mrobust.do

Replication in Sociology

Students in my graduate statistics course at Stanford asked 53 authors for a replication package

Selected articles they *admired and wanted to learn from*

Asked authors for the data and code for purposes of replication

Replication in Sociology

Students in my graduate statistics course at Stanford asked 53 authors for a replication package

Selected articles they *admired and wanted to learn from*

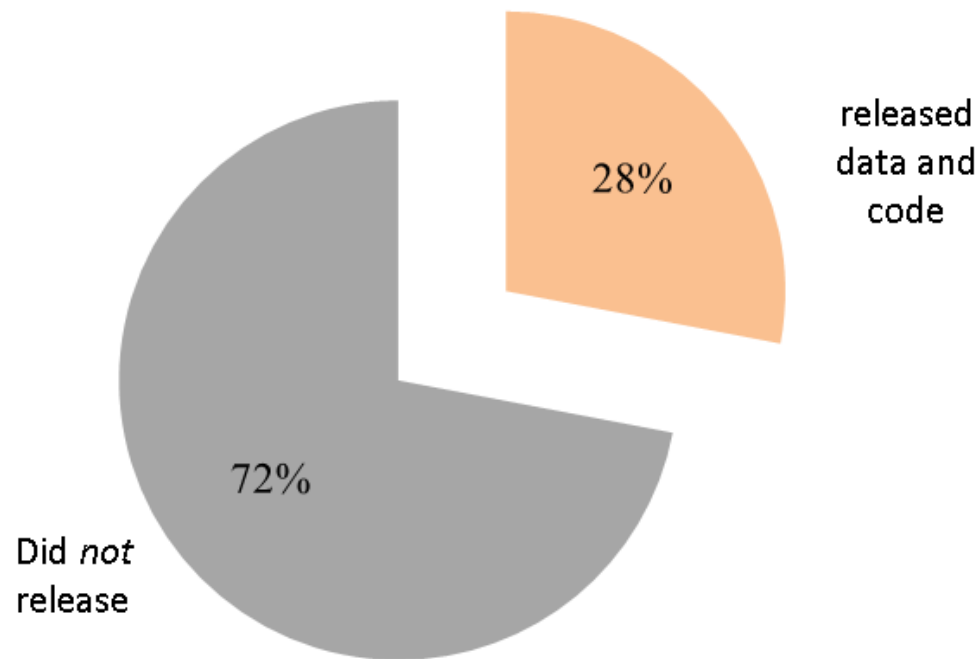
Asked authors for the data and code for purposes of replication

Guesses?

How many were willing and able to provide a replication package for their published article?

Replication in Sociology

Response to Replication Request (N = 53)



Note: For replication and transparency, [a blinded copy of the data is available on-line](#). Each author's identity is blinded, but the journal name, year of publication, and response code is available. Half of the requests addressed articles in the top three journals, and more than half were published in the last three years.

Replication in Sociology

Figure 1: Illustrative Quotes from Student Correspondence with Authors:

Positive:

1. “Here is the data file and Stata .do file to reproduce [the] Tables.... Let me know if you have any questions.”
2. “[Attached are] data and *R* code that does all regression models in the paper. Assuming that you know *R*, you could literally redo the entire paper in a few minutes.”

Replication in Sociology

Figure 1: Illustrative Quotes from Student Correspondence with Authors:

Positive:

1. “Here is the data file and Stata .do file to reproduce [the] Tables.... Let me know if you have any questions.”
2. “[Attached are] data and *R* code that does all regression models in the paper. Assuming that you know *R*, you could literally redo the entire paper in a few minutes.”

Negative:

3. “While I applaud your efforts to replicate my research, the best guidance I can offer is that the details about the data and analysis strategies are in the paper.”

Replication in Sociology

Figure 1: Illustrative Quotes from Student Correspondence with Authors:

Positive:

1. “Here is the data file and Stata .do file to reproduce [the] Tables.... Let me know if you have any questions.”
2. “[Attached are] data and *R* code that does all regression models in the paper. Assuming that you know *R*, you could literally redo the entire paper in a few minutes.”

Negative:

3. “While I applaud your efforts to replicate my research, the best guidance I can offer is that the details about the data and analysis strategies are in the paper.”
4. “I don’t keep or produce ‘replication packages’... Data takes a significant amount of human capital and financial resources, and serves as a barrier-to-entry against other researchers... they can do it themselves.”

Online Reactions

“ ‘Read the god damned paper and figure it out yourself’ is not an unreasonable response... the lazy motherf**kers who use the available codes are only going to find the same sh*t that the original author found”.

Online Reactions

“ ‘Read the god damned paper and figure it out yourself’ is not an unreasonable response... the lazy motherf**kers who use the available codes are only going to find the same sh*t that the original author found”.

Not having a replication package “was maybe fine in the 1990s... but now is a sign of somebody who is more slob than scientist and whose work should be evaluated accordingly.”

Online Reactions

“ ‘Read the god damned paper and figure it out yourself’ is not an unreasonable response... the lazy motherf**kers who use the available codes are only going to find the same sh*t that the original author found”.

Not having a replication package “was maybe fine in the 1990s... but now is a sign of somebody who is more slob than scientist and whose work should be evaluated accordingly.”

“Why don’t you publish the list of these people [who would not release their code]?”

Layout from here:

- 1) The downsides of transparency
- 2) The threat to professional self-governance of science
- 3) How to raise the standards of transparency
- 4) Journals – part 1: Leviathan “Law and Order” Econ – strict obligation

rewarding more transparency: Badges for



The Sorrows of Transparency

Reasons why authors *do not* want to release replication packages:

The Sorrows of Transparency

Reasons why authors *do not* want to release replication packages:

- (1) Replication packages require additional time and effort to prepare;

The Sorrows of Transparency

Reasons why authors *do not* want to release replication packages:

- (1) Replication packages require additional time and effort to prepare;
- (2) Other researchers can use the code to more quickly master the data and/or methodology, increasing the productivity of others but not oneself;

The Sorrows of Transparency

Reasons why authors *do not* want to release replication packages:

- (1) Replication packages require additional time and effort to prepare;
- (2) Other researchers can use the code to more quickly master the data and/or methodology, increasing the productivity of others but not oneself;
- (3) Readers may catch errors and mistakes in implementation that are not visible in the paper itself;

The Sorrows of Transparency

Reasons why authors *do not* want to release replication packages:

- (1) Replication packages require additional time and effort to prepare;
- (2) Other researchers can use the code to more quickly master the data and/or methodology, increasing the productivity of others but not oneself;
- (3) Readers may catch errors and mistakes in implementation that are not visible in the paper itself;
- (4) Critics may be able to quickly overturn the results using an alternative analysis that the author did not think of.

The Sorrows of Transparency

Reasons why authors *do not* want to release replication packages:

- (1) Replication packages require additional time and effort to prepare;
- (2) Other researchers can use the code to more quickly master the data and/or methodology, increasing the productivity of others but not oneself;
- (3) Readers may catch errors and mistakes in implementation that are not visible in the paper itself;
- (4) Critics may be able to quickly overturn the results using an alternative analysis that the author did not think of.

This list of drawbacks makes it clear that transparency mostly benefits other people, not the individual researcher.

What Way(s) Forward?

- (1) Enforce required transparency as a condition of publication**
 - as in economics**

What Way(s) Forward?

- (1) Enforce required transparency as a condition of publication**
 - as in economics**

What Way(s) Forward?

- (1) Enforce required transparency as a condition of publication**
 - as in economics**

- (2) Encourage transparency with a simple, effective “nudge”**
 - as in psychology**



**AMERICAN
ECONOMIC
ASSOCIATION**

Data Availability Policy

It is the policy of the American Economic Association to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication.

Policy applies to all official AEA journals.

Since adopted by all top journals in economics

No Price Like Home: Global House Prices, 1870-2012

Katharina Knoll

Moritz Schularick

Thomas Steger

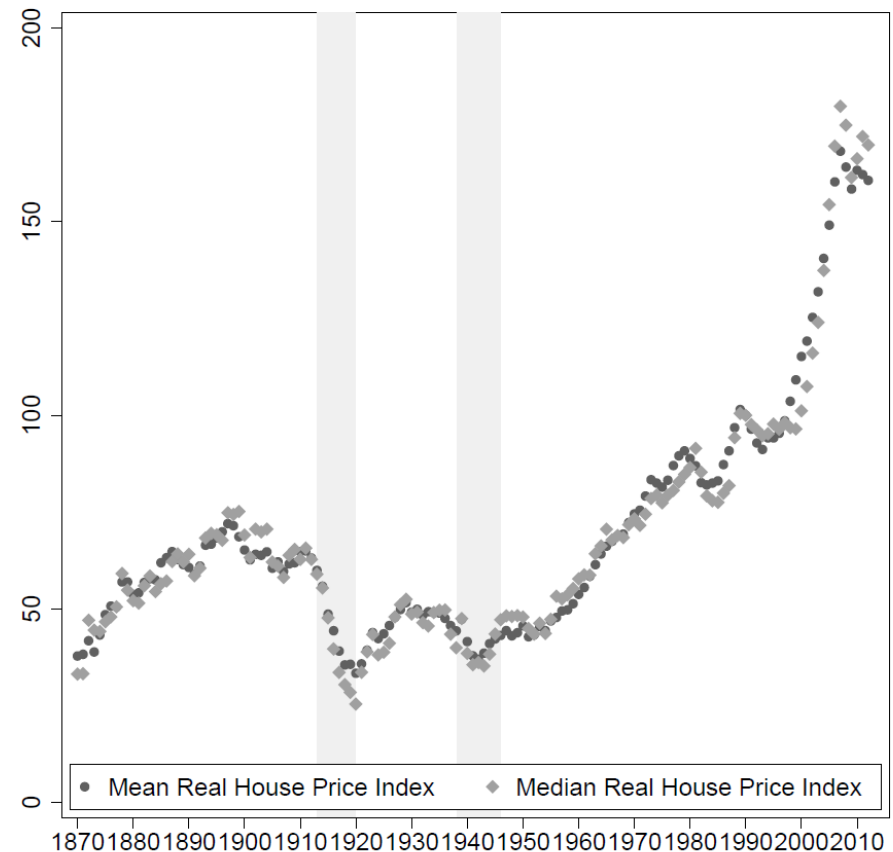
AMERICAN ECONOMIC REVIEW
VOL. 107, NO. 2, FEBRUARY 2017
(pp. 331-53)

[Download Full Text PDF](#)

Additional Materials

[Data Set \(2.36 MB\)](#)

[Online Appendix \(2.47 MB\)](#)



Nudges to reward more transparency: Badges for openness

In other social sciences, the data availability policy in economics faces too much resistance.

It is not currently feasible for journals to require openness and transparency.

But some people are already making replication packages available

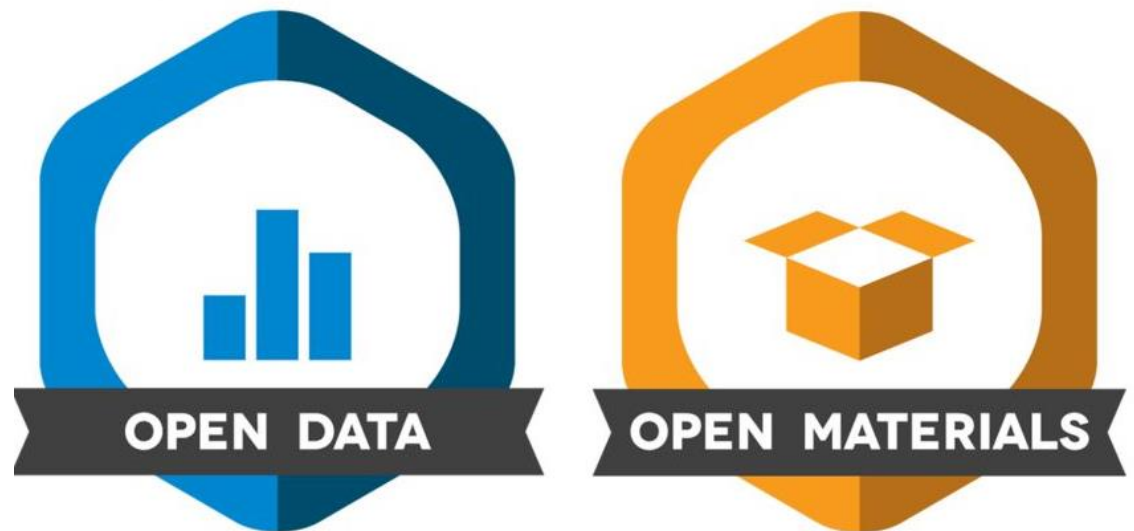
Nudges to reward more transparency:

Badges for openness

In other social sciences, the data availability policy in economics faces too much resistance.

It is not currently feasible for journals to require openness and transparency.

But some people are already making replication packages available



The Belief in a Favorable Future



Todd Rogers¹, Don A. Moore², and
Michael I. Norton³

¹Harvard Kennedy School, Harvard University; ²Haas School of Business, University of California, Berkeley; and ³Harvard Business School, Harvard University

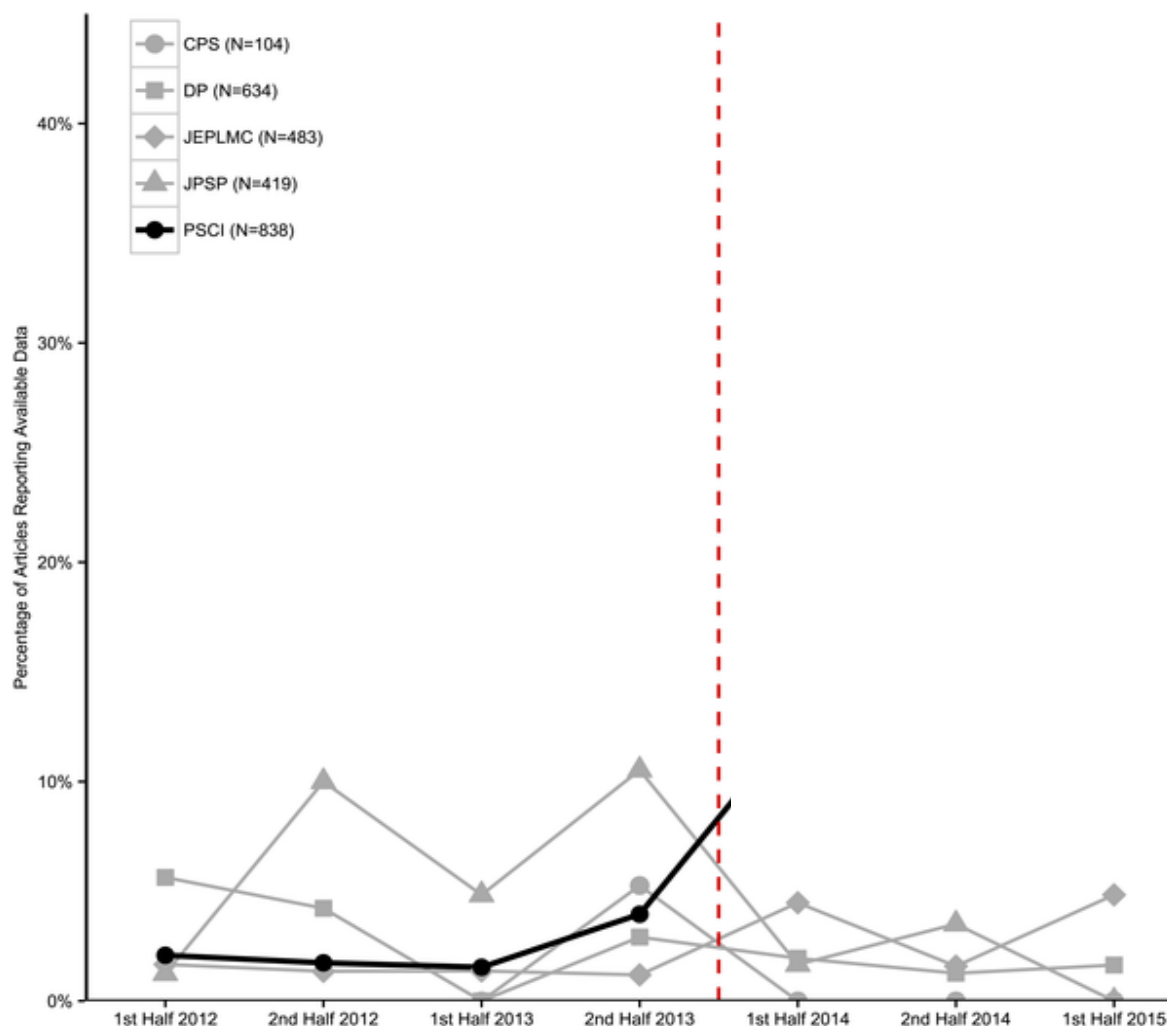
Psychological Science
2017, Vol. 28(9) 1290–1301
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797617706706
www.psychologicalscience.org/PS



Abstract

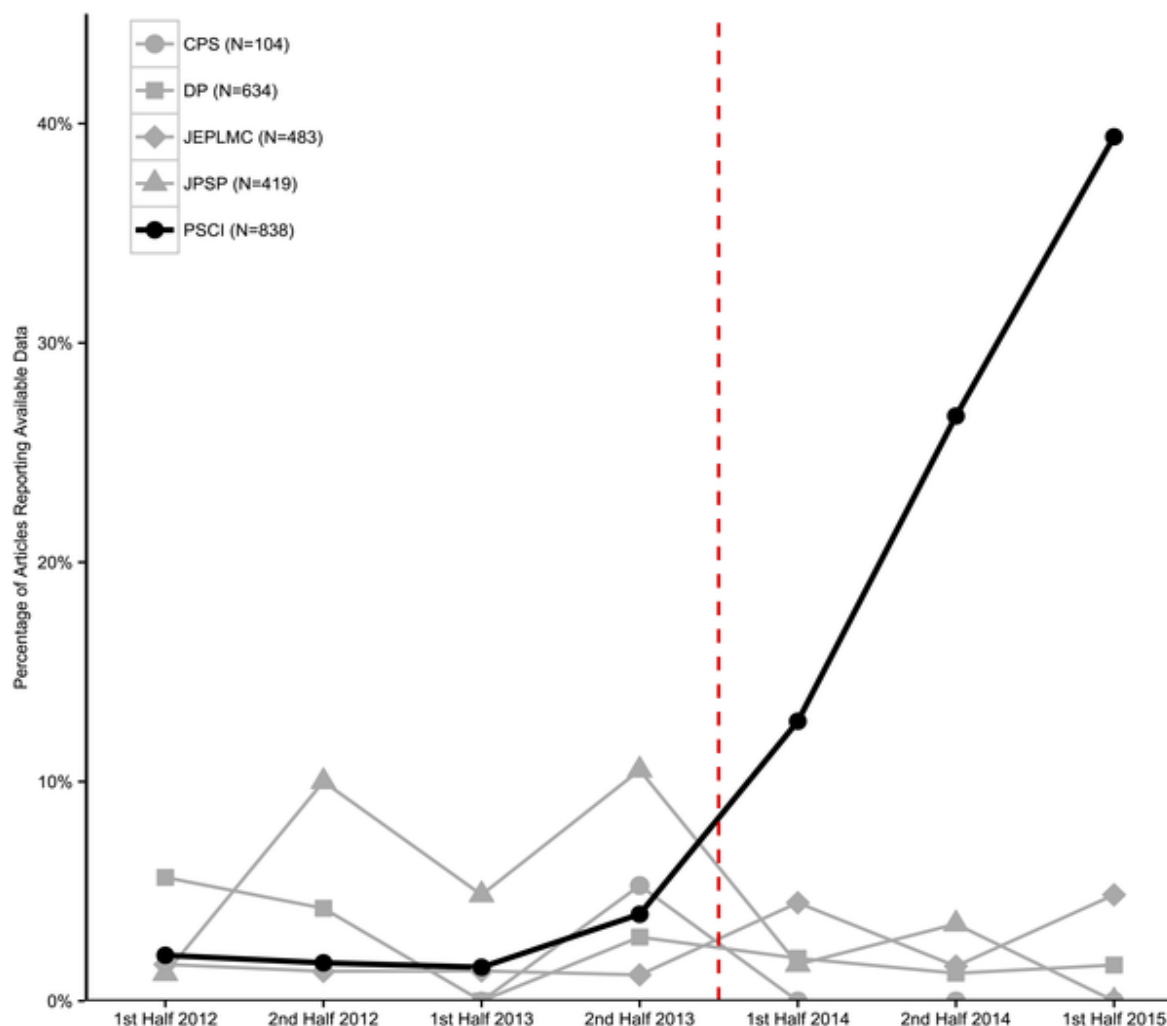
People believe that future others' preferences and beliefs will change to align with their own. People holding a particular view (e.g., support of President Trump) are more likely to believe that future others will share their view than to believe that future others will have an opposing view (e.g., opposition to President Trump). Six studies demonstrated this *belief in a favorable future* (BFF) for political views, scientific beliefs, and entertainment and product preferences. BFF is greater in magnitude than the tendency to believe that current others share one's views (false-consensus effect), arises across cultures, is distinct from general optimism, is strongest when people perceive their views as being objective rather than subjective, and can affect (but is distinct from) beliefs about favorable future policy changes. A lab experiment involving monetary bets on the future popularity of politicians and a field experiment involving political donations ($N = 660,542$) demonstrated that BFF can influence people's behavior today.

Open Data in Psychology Journals, before and after “badges” introduced



Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, et al. (2016) Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology* 14(5): e1002456. <https://doi.org/10.1371/journal.pbio.1002456>
<http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002456>

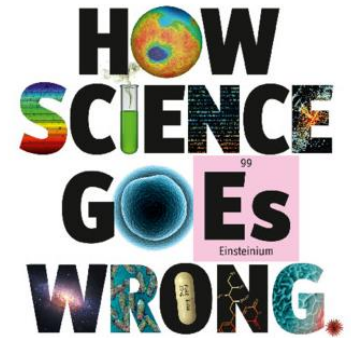
Open Data in Psychology Journals, before and after “badges” introduced



Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, et al. (2016) Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology* 14(5): e1002456. <https://doi.org/10.1371/journal.pbio.1002456>
<http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002456>

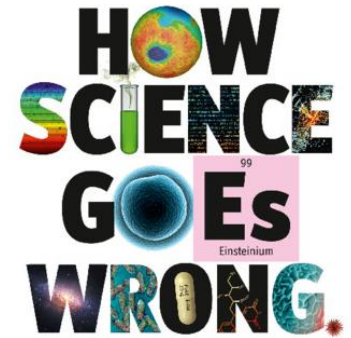
Conclusion:

We are facing today is declining public trust in science.



Conclusion:

We are facing today is declining public trust in science.



Return to one of the founding principles of science:

Nullius in Verba – take no one's word for it

see for yourself!



Thank you!

Contact:

Cristobal Young

cristobal.young@stanford.edu

Website:

www.cristobalyoung.com



See for yourself!



Drawback of administrative data:

Company data cannot be made publically available – no company would agree to that.

This means no one can ever go back and double check Goldberg et al's analysis. (though they should be able to examine the statistical code they used)

This is a problem for transparency. Econ journals today require data sharing for all published papers, unless the data cannot be legally shared (which is increasingly common).

At least one major journal - Review of Economics and Statistics – will not accept or publish articles that use strictly private data. If a serious scholar cannot apply for and reasonably expect data access, the article cannot be published in RES.

With IRS data, it is *possible* to apply for data access to replicate & extend Young et al. But an author could only get access if an IRS analyst personally wanted to pursue the project (for publication prospects).