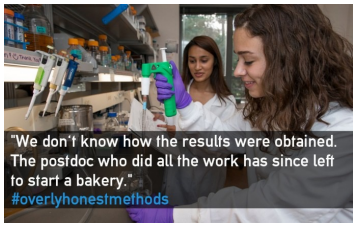
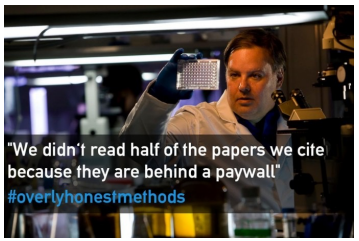
COGS 105

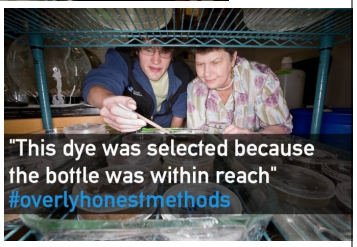
Research Methods for Cognitive Scientists

Week 5, Class 2:
Behavioral Methods IV: Research Ethics

Cynical Science Memes…

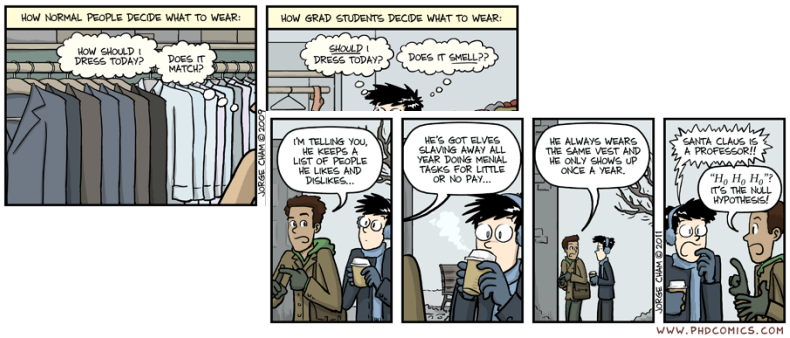






Cynical Science Comics…

- Famous "Ph.D. Comics" (a favorite of graduate students everywhere…).



Cognitive **Science**

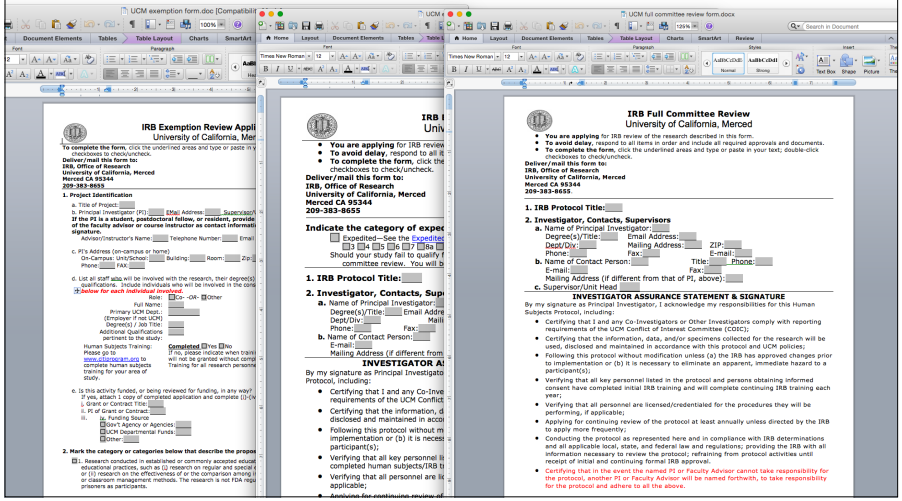
Research **Ethics**

- Several ethical issues
  - Safety and fairness to participants in the study
  - Honesty and accuracy of scientific practices
  - Accuracy of reporting to the community
  - Data sharing and openness to the community

Safety Issues

- **IRB** = Institutional Review Board
  - Each university has one (sometimes more, for different types of research)
  - Tracks **compliance** with federal law; oversees research by all at a university whose work might impact people and animals
  - Federal law: Health and Human Services (HHS) and FDA Protection of Human Subjects Regulations
- Researchers **submit IRB applications** for approval; doing your research before approval can lead to punishment (e.g., ban from research or worse).

UC Merced's IRB


IRB Exemption Review Appli
University of California, Mer

IRB
Univ

IRB Full Committee Review
University of California, Merced

"Exemption"

- Most basic cognitive science research, such as RT experiments, is deemed "exempt" from any real risk. These experiments that get permission in an expedited fashion.
  - Risk? "You might become slightly bored."

E.g.: Exemption

Types of IRB Review

Exempt Research

Although the category is called "exempt," this type of research does require IRB review and the determination of exemption must be made by the IRB.

The exempt determination process is much less rigorous than an expedited or full-committee review. To qualify, research must fall into one or more of six federally defined exempt categories.

These categories present the lowest amount of risk to potential subjects because, generally speaking, they involve either collection of anonymous or public data, or conduct of the least potentially harmful research experiments. Some examples of exempt research are:

- anonymous surveys or interviews
- passive observation of public behavior without collection of subject identifiers
- retrospective chart reviews
- analyses of discarded pathological specimens without patient identifiers

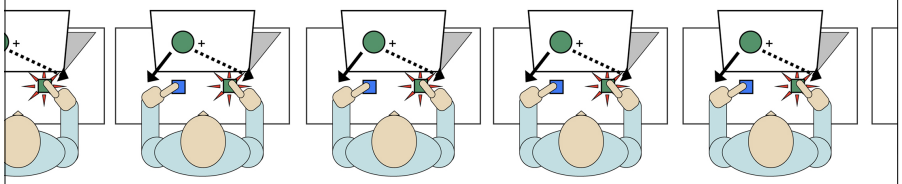
http://rci.ucmerced.edu/irb/researchers/types-review

Research **Ethics**

- Several ethical issues
  - Safety and fairness to participants in the study
  - Honesty and accuracy of scientific practices
  - Accuracy of reporting to the community
  - Data sharing and openness to the community

Replication

- If we setup the same conditions again and run the study, the result should essentially be the same (allowing for statistical variability that is expected).
  - Why would it **not** be the same — results are different — after trying to replicate it?



Scientific Practices

- When we publish a paper on our newest awesome finding, it is the end result of a **long series of scientific decisions** we make.
  - Task design, stimulus design, subject recruitment, data management and cleaning, data analysis, and so on…
- Typically, it is expected that **we report our work in such detail that others could replicate it**, but this is difficult to achieve in a single paper that often has length restrictions.

"Experimenter Degrees of Freedom"

- But we face an important problem, related to last class, too: Scientists themselves can be impacted by their wants, needs, desires, etc. ***careers***.
  - This means scientists might *can* impact subtle decisions about the study that could **bias in favor** of getting a positive outcome of the study.
  - These decisions can be called "**experimenter degrees of freedom**," and they are often unaccounted for after the study is done (Simmons et al., 2011).

Classic Example

- When have you run enough participants to **stop** and **report** your work to the scientific community?
- **Optimal strategy**: Sit down, look at past findings, and carefully work out how many subjects you are going to run **in advance** (general rule: the more, the better). "**Power analysis**."
- **Bad strategy**: Think of a number, run that number, do a quick check and then "hey, looks promising, let's run some more." Iterate. This is using your "experimenter degrees of freedom" — you are biased to stop when you get your effect! Sometimes called "**p-hacking**."

General Article

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
XX(X) 1–8

DOI: 10.1177/0956797611417632
http://pss.sagepub.com
SAGE

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings (≤ .05), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Keywords

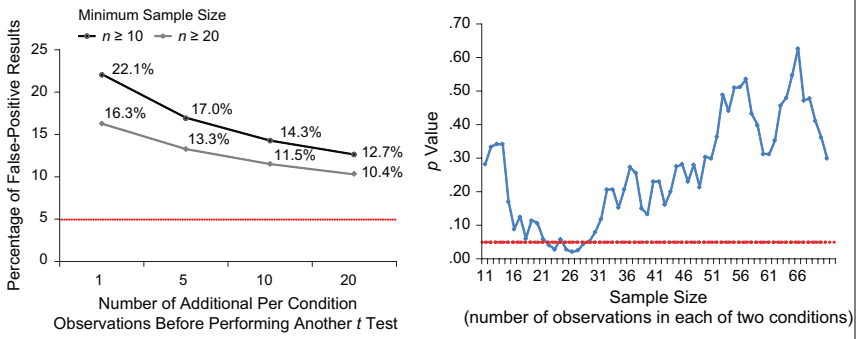
methodology, motivated reasoning, publication, disclosure

Received 3/17/11; Revision accepted 5/23/11

Our job as scientists is to discover truths about the world. We generate hypotheses, collect data, and examine whether or not the data are consistent with those hypotheses. Although we

Which control variables should be considered? Should specific measures be combined or transformed or both?

Example from Simmons et al.



Running more subjects iteratively and then checking can lead to false positives. The data above are from computer simulations (where there is no effect) showing you can get high % of false positives under this strategy.

Research **Ethics**

- Several ethical issues
  - Safety and fairness to participants in the study
  - Honesty and accuracy of scientific practices
  - Accuracy of reporting to the community
  - Data sharing and openness to the community

"Publication Bias"

- Scientists and professors are stressed out.
  - In addition to teaching, they spend countless hours contributing to their scientific community and advising graduate students and doing research.
- Unfortunately, that means that succumbing to stress is possible — all people can be lazy, they can be biased to publish crappy results, not driven to carefully check our descriptions in papers, etc.
  - **Problematic incentive structure** of academic science?

"Publication Bias"

- Here's an example of not reporting everything you've run to your community in order to bias towards publication.
- **Run 20 studies**. By chance, how many are probably going to have *something* in them that is significant, according to a p-value?
  - **1 out of 20**; p = .05 or less.
- **Publication bias**: Tuck the 19 studies "in your file drawer" and publish just that 1 study.

Example?

Psychon Bull Rev (2012) 19:151–156
DOI 10.3758/s13423-012-0227-9

BRIEF REPORT

Too good to be true: Publication bias in two prominent studies from experimental psychology

Gregory Francis

Published online: 15 February 2012

Abstract Empirical replication has long been considered the final arbiter of phenomena in science, but replication is undermined when there is evidence for publication bias. Evidence for publication bias in a set of experiments can be found when the observed number of rejections of the null hypothesis exceeds the expected number of rejections. Application of this test reveals evidence of publication bias in two prominent investigations from experimental psychology that have pur-

findings is essentially the same error as filtering out subjects who do not behave in a desired way. Even well-designed studies can be rendered scientifically useless if other studies are done poorly and publication bias contaminates the set.

Here, publication bias is investigated in two prominent sets of results from experimental psychology. These studies have attracted widespread attention in both academic and nonacademic reports, because they appear to challenge the estab-

Bem (2011)

- Francis challenges the following paper on the grounds that it probably does not report all of the experiments that were conducted.

Journal of Personality and Social Psychology

0022-3514/11/$12.00 DOI: 10.1037/a0021524

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some

Power and Chance

- If you have a series of low-powered studies, with a weak or striking effect, several of these studies shouldn't pan out. So Francis argues:

> Perhaps the most striking property of Bem (2011) is that nine out of 10 described investigations rejected the null hypothesis, thereby indicating evidence for psi. For many scientists, replication of an effect across multiple experiments provides compelling evidence, but this interpretation is misguided, because it does not consider the statistical power of the experiments. If all of the experiments have high power (the probability of rejecting the null hypothesis when it is false), multiple experiments that reject the null hypothesis would indeed be strong evidence for an effect. However, if the experiments have low or moderate power, then even if the effect were real, one would expect to frequently not reject the null hypothesis.

Community Orientation

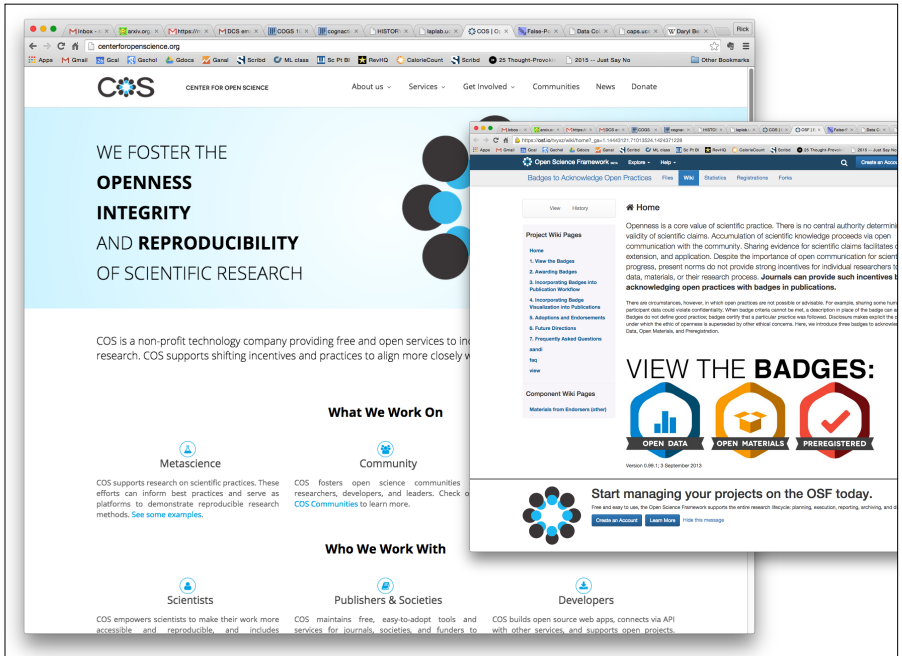
- It is important to remember that any accusation of publication bias or "p-hacking" must be taken with standard assumptions of **innocence until proven guilty** — maybe the experimenter got lucky, maybe a power analysis would have come to same conclusion, etc.
- A **focused awareness** of these potential biases is a very important feature for a scientist to have.
  - You do not want your findings **questioned in the community**; if the replication does not work out, your paper will be consigned to the **trash bin of false findings**; even worse, it could **damage your reputation**.

Research **Ethics**

- Several ethical issues
  - Safety and fairness to participants in the study
  - Honesty and accuracy of scientific practices
  - Accuracy of reporting to the community
  - Data sharing and openness to the community

Replication Agenda

- A community of researchers has been developing **open science practices**.
- These biases can be avoided if we have **better research practices**:
  - **Be clear on the front end** how you will do your study
  - **Share your results** and data analysis scripts
  - **Share your materials** so your work can be replicated



Reproducibility Project

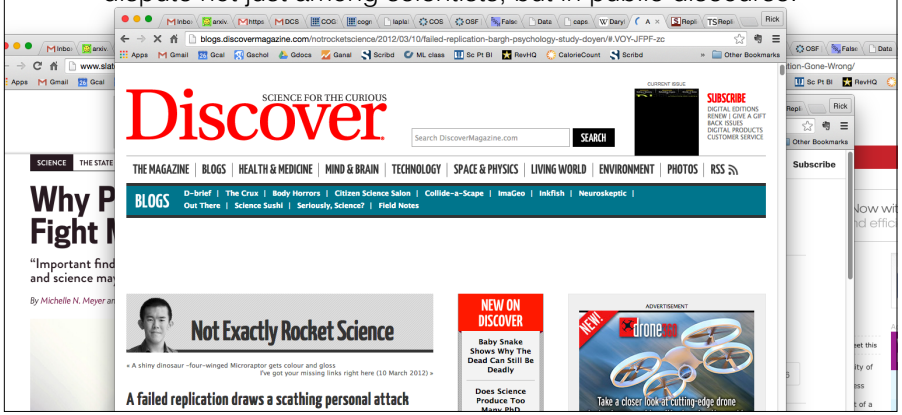
- The “Many Labs” project (the required reading; just 10 pages).
  - Also a broader “Reproducibility Project”: Replicate *all papers* from 2008 in *Psychological Science*, *Journal of Experimental Psychology: Learning, Memory & Cognition*, and *Journal of Personality and Social Psychology*.
    - Over a hundred experiments! Still ongoing.

Controversy

- This incredibly important project has led to some dispute in the field.
  - **How precise** should replications be? Could direct replications be difficult if effects were found at a different time, different cultural context, etc.?
- "**Conceptual replication**" vs. "**direct replication**"
- These have also induced "human issues" — sensationalist tweets induce anger in original authors whose work may not have replicated (or, at least, perfectly replicated).
  - Important ethical dimension: **Fairness to fellow researchers** in considering these replication projects, also important too.

Whoa

There has been some emphasis in replicating social priming effects because they seem so striking. The result has been considerable dispute not just among scientists, but in public discourse.



Being Fair

"Headcon" Event

There has been dispute about the fairness of some of the replication practices; do they give the original authors enough air time to address issues?

http://edge.org/conversation/headcon-14

The Obvious Truth Here…

- **Replication is crucial**, and the agenda of the Open Science Framework / Center for Open Science will be a **leading force in the change** not just in psychology and cognitive science, but perhaps even throughout science…
  - And the specific practices we employ to develop replications and tests of other peoples' work should also be **fair and collegial**, in order to further **strengthen the openness vibe**.
  - And, importantly: We must **allow the possibility that our work will prove to be wrong**. If we do not allow for this, and we allow ourselves to be biased against this possibility, then we are not scientists. (Feynman's "don't fool ourselves.)

"Many Labs" Project

Replication

Investigating Variation in Replicability

A "Many Labs" Replication Project

Richard A. Klein,[1] Kate A. Ratliff,[1] Michelangelo Vianello,[2] Reginald B. Adams Jr.,[3] Štěpán Bahník,[4] Michael J. Bernstein,[5] Konrad Bocian,[6] Mark J. Brandt,[7] Beach Brooks,[1] Claudia Chloe Brumbaugh,[8] Zeynep Cemalcilar,[9] Jesse Chandler,[10,36] Winnee Cheong,[11] William E. Davis,[12] Thierry Devos,[13] Matthew Eisner,[10] Natalia Frankowska,[6] David Furrow,[15] Elisa Maria Galliani,[2] Fred Hasselman,[16,37] Joshua A. Hicks,[12] James F. Hovermale,[17] S. Jane Hunt,[18] Jeffrey R. Huntsinger,[19] Hans IJzerman,[7] Melissa-Sue John,[20] Jennifer A. Joy-Gaba,[17] Heather Barry Kappes,[21] Lacy E. Krueger,[18] Jaime Kurtz,[22] Carmel A. Levitan,[23] Robyn K. Mallett,[19] Wendy L. Morris,[24] Anthony J. Nelson,[3] Jason A. Nier,[25] Grant Packard,[26] Ronaldo Pilati,[27] Abraham M. Rutchick,[28] Kathleen Schmidt,[29] Jeanine L. Skorinko,[20] Robert Smith,[14] Troy G. Steiner,[3] Justin Storbeck,[8] Lyn M. Van Swol,[30] Donna Thompson,[15] A. E. van 't Veer,[7] Leigh Ann Vaughn,[31] Marek Vranka,[32] Aaron L. Wichman,[33] Julie A. Woodzicka,[34] and Brian A. Nosek[29,35]

[1]University of Florida, Gainesville, FL, USA, [2]University of Padua, Italy, [3]The Pennsylvania State University, University Park, PA, USA, [4]University of Würzburg, Germany, [5]Pennsylvania State University Abington, PA, USA, [6]University of Social Sciences

Table 1. Data collection sites

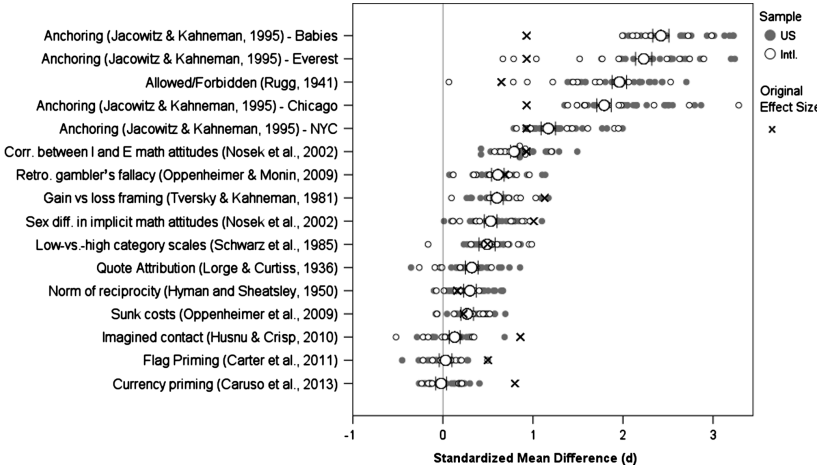
Site identifier	Location	*N*	Online (O) or laboratory (L)	US or international (I)
Abington	Penn State Abington, Abington, PA	84	L	US
Brasilia	University of Brasilia, Brasilia, Brazil	120	L	I
Charles	Charles University, Prague, Czech Republic	84	L	I
Conncoll	Connecticut College, New London, CT	95	L	US
CSUN	California State University, Northridge, LA, CA	96	O	US
Help	HELP University, Malaysia	102	L	I
Ithaca	Ithaca College, Ithaca, NY	90	L	US
JMU	James Madison University, Harrisonburg, VA	174	O	US
KU	Koç University, Istanbul, Turkey	113	O	I
Laurier	Wilfrid Laurier University, Waterloo, Ontario, Canada	112	L	I
LSE	London School of Economics and Political Science, London, UK	277	L	I
Luc	Loyola University Chicago, Chicago, IL	146	L	US
McDaniel	McDaniel College, Westminster, MD	98	O	US
MSVU	Mount Saint Vincent University, Halifax, Nova Scotia, Canada	85	L	I
MTURK	Amazon Mechanical Turk (US workers only)	1,000	O	US
OSU	Ohio State University, Columbus, OH	107	L	US
Oxy	Occidental College, LA, CA	123	L	US
PI	Project Implicit Volunteers (US citizens/residents only)	1,329	O	US
PSU	Penn State University, University Park, PA	95	L	US
QCCUNY	Queens College, City University of New York, NY	103	L	US
QCCUNY2	Queens College, City University of New York, NY	86	L	US
SDSU	SDSU, San Diego, CA	162	L	US
SWPS	University of Social Sciences and Humanities Campus Sopot, Sopot, Poland	79	L	I
SWPSON	Volunteers visiting www.badania.net	169	O	I
TAMU	Texas A&M University, College Station, TX	187	L	US
TAMUC	Texas A&M University-Commerce, Commerce, TX	87	L	US
TAMUON	Texas A&M University, College Station, TX (Online participants)	225	O	US
Tilburg	Tilburg University, Tilburg, Netherlands	80	L	I
UFL	University of Florida, Gainesville, FL	127	L	US
UNIPD	University of Padua, Padua, Italy	144	O	I
UVA	University of Virginia, Charlottesville, VA	81	L	US
VCU	VCU, Richmond, VA	108	L	US
Wisc	University of Wisconsin-Madison, Madison, WI	96	L	US
WKU	Western Kentucky University, Bowling Green, KY	103	L	US
WL	Washington & Lee University, Lexington, VA	90	L	US
WPI	Worcester Polytechnic Institute, Worcester, MA	87	L	US

2 Examples

9. *Flag Priming (Carter, Ferguson, & Hassin, 2011; Study 2)*. The American flag is a powerful symbol in American culture. Carter et al. (2011) examined how subtle exposure to the flag may increase conservatism among US participants. Participants were presented

liberal. (1 = *Liberal*, 7 = *Conservative*).

10. *Currency priming (Caruso, Vohs, Baxter, & Waytz, 2013)*. Money is a powerful symbol. Caruso et al. (2013) provide evidence that merely exposing participants to money increases their endorsement of the current social system. Participants were first pre-



Cohen's d = effect size of the replication…

Next Week

- **Review and discussion on Tuesday: Study guide is on the website**.
- Sections next week: Starting “Big Data” and computer-science-related methods.
  - Hands-on exploration of books and brains!