Lecture 4: PROBLEMS OF SLIPPERY SCIENCE

A-- The Slippery Slope

Would you go down the slippery slope?

1)Your experiment worked but an aberrant observation messes things up. If you found a reason to throw it out, you would have a powerful article for this years' job market. If not, goodbye job. What do you do?

2)You find a mistake in your recent published paper that makes result shaky. Should you retract and declare it was your mistake? Blame your post-doc? Keep shut and hope no one else notices the error.

3)Your team has produced remarkable outcome that will be front page in Science and Nature. Competitors are close on your tail. Should you spend your budget replicating your striking result or rush to publication?

4)You can add a co-author to your paper and get rewarded with a $5k gift? … a greater chance of acceptance at top journal (the co-author is spouse of editor)? Should you do this? What if the offer is $15K … guaranteed acceptance?

5)You cite the editor of the journal to which you submit your paper and all the other authors in your field who published in the journal, to increase chance you will get a favorable reading of your paper.

6)You notice a fellow researcher treating his/her data suspiciously. Should you confront the person? Report them to someone higher up in the scientific world? Ignore the incident and keep doing your own work?

7)You read an unpublished article by a scientific neophyte working in the jungles of Sumatra that solves the problem that has occupied your life. Do you claim you knew the solution and publish a "joint" article and then rush out a book to assure your priority? Declare the other person has the answer and advise them how to scoop you?

8) XYZ Corporation has given you a $2M dollar grant to study the health value of taking XYZ supplements. You do experiments and find no results for big sample of people but note that one group – teenage boys – has a beneficial effect. Do you declare "supplement fails" or instead report that it works for teen boys?

9)Your drug firm finds that the Ponce de Leon pill makes middle aged women feel younger but has no effect of men. If you pool the middle aged women with men, still get significant effect. Twice the population to sell the drug to … twice the profits … Do you bring the whole sample for FDA trial or just the group on which it worked?

10) You are invited to join a group of researchers who give highly favorable peer reviews to papers of the group. All you have to do is say papers of members should be accepted and they will do the same for you. This will increase your chance of getting published and promoted.

The slippery problems of science are now widely recognized

Figure errors, sloppy science, and fraud: keeping eyes on your data

Corinne L. Williams, … , Arturo Casadevall, Sarah Jackson

J Clin Invest. 2019;129(5):1805-1807. https://doi.org/10.1172/JCI128380.

Editorial

Recent reports suggest that there has been an increase in the number of retractions and corrections of published articles due to post-publication detection of problematic data. Moreover, fraudulent data and sloppy science have long-term effects on the scientific literature and subsequent projects based on false and unreproducible claims. At the *JCI*, we have introduced several data screening checks for manuscripts prior to acceptance in an attempt to reduce the number of post-publication corrections and retractions, with the ultimate goal of increasing confidence in the papers we publish.

Between July 1, 2018 and February 5, 2019, we screened 200 papers that were on a clear path toward acceptance for publication. These papers are typically revised manuscripts that have only minor points leftto address prior to acceptance … During the course of our study, 28.5% (57 of 200) of the papers screened wereflagged for issues with statistical tests, 21% (42 of 200) of papers had some issuewith the blots, and 27.5% (55 of 200) of papers had issues with images

Magnitude and nature of falsifying/other bad practices

How can we estimate what may be tip of iceberg? Will be a bit like Drake's laws for probability of intelligent life on distant planet. Some statistical calculations but based on problematic guesses. Many small slippery actions unlikely to be found but then maybe they do not matter.

One way to estimate magnitude is to **ask people.**

Daniele Fanelli PLOSONE 2009 How Many Scientists Fabricate and Falsify Research? A Systematic Review **and Meta-Analysis of Surveys** that asked scientists directly whether they have committed or know of a colleague who committed research misconduct. To standardize outcomes, the number of respondents who recalled at least one incident of misconduct was calculated for each question, and the analysis was limited to behavior that distorts scientific knowledge:fabrication, falsification, "cooking" of data, etc… The final sample consisted of 21 surveys in the systematic review, and 18 in meta-analysis**. A pooled weighted average of 1.97% (N = 7, 95%CI: 0.86–4.45) of scientists admitted to have fabricated, falsified or modified data or results at least once; 33.7% admitted other questionable research practices.** In surveys asking about behavior of colleagues, admission rates were **14.12% (N = 12, 95% CI: 9.91–19.72) for falsification, and up to 72%** for other questionable practices.

Self reports surveys, surveys using the words "falsification" or "fabrication", and mailed surveys yielded lower percentages of misconduct. When these factors were controlled for, misconduct was reported more frequently by medical/pharmacological researchers than others. Considering that these surveys ask sensitive questions and have other limitations, likely that this is a conservative estimate of the true prevalence of scientific misconduct.

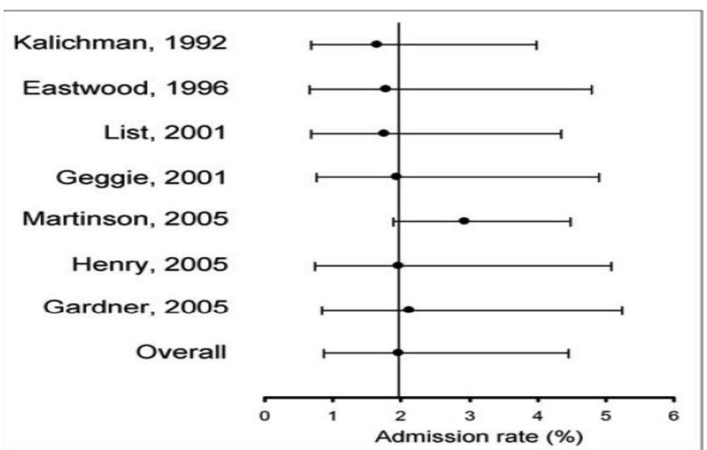


Figure 5. Sensitivity analysis of admission rates of data fabrication, falsification and alteration in self reports. Plots show the weighted pooled estimate and 95% confidence interval obtained when the corresponding study was left out of the analysis. doi:10.1371/journal.pone.0005738.g005

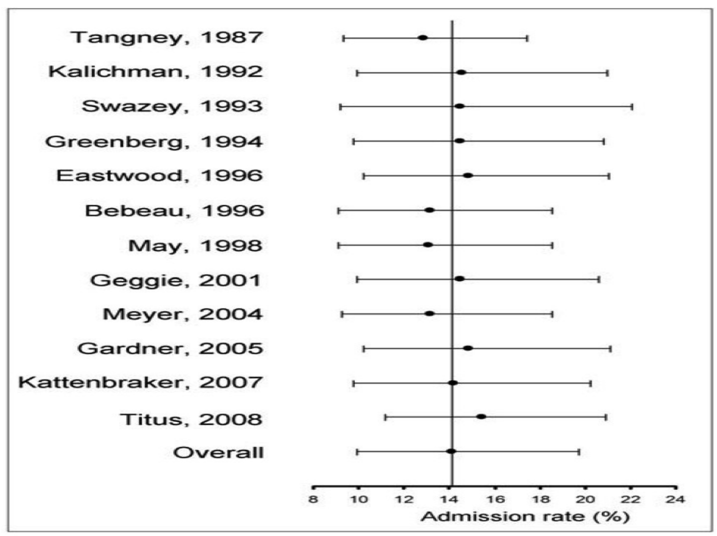


Figure 6. Sensitivity analysis of admission rates of data fabrication, falsification and alteration in non-self reports.

Leslie K. John , George Loewenstein, and Drazen Prelec " Measuring the Prevalence of Questionable Research Practices (QRPs) With Incentives for Truth Telling" Psychological Science 23(5) 524–532 2012 electronic survey to 5,964 academic psychologists with 2155 –36% response rate at major U.S. Universities.

Respondents asked whether they had personally engaged in each of 10 QRPs; their view of prevalence among others; and among those who had engaged in QRP the % who admitted it. This gives bunch of estimates of the effects:

% of scientists who admit QRP(self-admission);

% they think do QRP (prevalence estimate)

% they think report QRP --- If take Self-admit by estimate of the truthfulness get another estimate of rate of QRP … if you think 90% report they do QRP adjust rate of self admit by 1/0.9 and get estimate of QPR. If persons are consistent – prevalence would equal average self-admit/expected truthfulness. Analysis based on averages, but can do it for each respondent –

They then add a truth-telling incentive by telling people they will get reward closer the average self-admission is to the predicted. Idea is people likely to understate self admit, so $$ should move them to raise the self-admit. Say I believe everyone cheats. If I say I do not cheat, self-admission will be less than the rate, and payment falls.

Table 1. Results of the Main Study: Mean Self-Admission Rates, Comparison of Self-Admission Rates Across Groups, and Mean Defensibility Ratings

Item	Self-admission rate (%)		Odds ratio (BTS/control)	Two-tailed *p* (likelihood ratio test)	Defensibility rating (across groups)
	Control group	BTS group			
1. In a paper, failing to report all of a study's dependent measures	63.4	66.5	1.14	.23	1.84 (0.39)
2. Deciding whether to collect more data after looking to see whether the results were significant	55.9	58.0	1.08	.46	1.79 (0.44)
3. In a paper, failing to report all of a study's conditions	27.7	27.4	0.98	.90	1.77 (0.49)
4. Stopping collecting data earlier than planned because one found the result that one had been looking for	15.6	22.5	1.57	.00	1.76 (0.48)
5. In a paper, "rounding off" a *p* value (e.g., reporting that a *p* value of .054 is less than .05)	22.0	23.3	1.07	.58	1.68 (0.57)
6. In a paper, selectively reporting studies that "worked"	45.8	50.0	1.18	.13	1.66 (0.53)
7. Deciding whether to exclude data after looking at the impact of doing so on the results	38.2	43.4	1.23	.06	1.61 (0.59)
8. In a paper, reporting an unexpected finding as having been predicted from the start	27.0	35.0	1.45	.00	1.50 (0.60)
9. In a paper, claiming that results are unaffected by demographic variables (e.g., gender) when one is actually unsure (or knows that they do)	3.0	4.5	1.52	.16	1.32 (0.60)
10. Falsifying data	0.6	1.7	2.75	.07	0.16 (0.38)

Note: Items are listed in decreasing order of rated defensibility. Respondents who admitted to having engaged in a given behavior were asked to rate whether they thought it was defensible to have done so (0 = *no*, 1 = *possibly*, and 2 = *yes*). Standard deviations are given in parentheses. BTS = Bayesian truth serum. Applying the Bonferroni correction for multiple comparisons, we adjusted the critical alpha level downward to .005 (i.e., .05/10 comparisons).

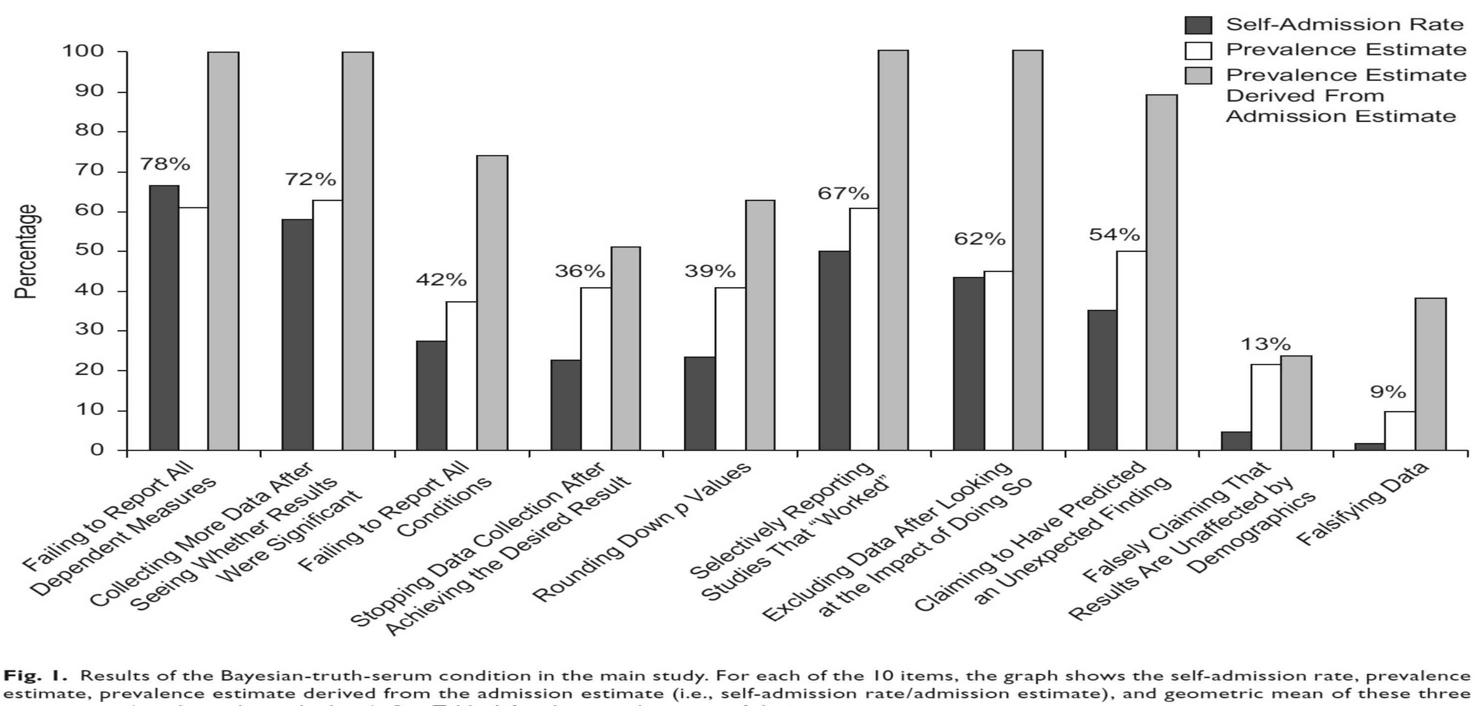


Fig. 1. Results of the Bayesian-truth-serum condition in the main study. For each of the 10 items, the graph shows the self-admission rate, prevalence estimate, prevalence estimate derived from the admission estimate (i.e., self-admission rate/admission estimate), and geometric mean of these three percentages (numbers above the bars). See Table 1 for the complete text of the items.

Andrew Wakefield’s fraudulent paper on vaccines and autism has been cited more than a thousand times. These researchers tried to figure out why.

Retraction Watch readers are no doubt familiar with one of the most consequential retractions of this century, namely that of the 1998 paper in *The Lancet* by Andrew Wakefield and others claiming a link between vaccines and autism. What they may also know is that the paper remains one of the most highly cited retracted articles of all time, as demonstrated by our leaderboard of such papers. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2755486

Harvard group retracts Nature paper … A group of researchers based at Harvard University have retracted an influential 2017 letter in *Nature* after a change in lab personnel led to the discovery of errors in the analysis. The article, “Microglia-dependent synapse loss in type I interferon-mediated lupus,” emerged from a collaboration including scientists at Harvard Medical School, the Rockefeller University in New York City, the University of Magdeburg, in Germany.The senior author of the research letter — which has been cited 75 times, earning it a highly cited designation from Clarivate Analytics’ Web of Science — was Michael C. Carroll, a prominent immunology researcherpediatric cancer specialist.Also on the list was Ronald Herbst, who at the time was vice president of research at MedImmune but has since left that company for another biotech firm. The first author was Allison Bialas, at the time a post-doc at Harvard. ***We realized that the data supporting the figures noted in the retraction could not be repeated*** *when a PhD student in the lab took over the project from the lead post-doc. No other retractions are planned.* Here’s the retraction notice:*In follow-up experiments to this Letter, we have been unable to replicate key aspects of the original results. Most importantly, the findings from behaviour studies and sequencing of microglia isolated from 564Igi autoimmune mice as shown in Figs. 1a, b, d and 3a, b are not substantiated upon further analysis of the original data. The authors therefore wish to retract the Letter. We deeply regret this error and apologize to our scientific colleagues.***‘It’s inspired my group’** The article was considered important enough to merit a News & Views piece in *Nature* when it appeared. It also garnered some media attention. We also found evidence that it influenced at least one other scientist. Edward Vital,a lupus researcher in England, tweeted that he was impressed by the results and hoped to apply them in his own work:

Disgraced Korea scholar, formerly of Columbia, loses paper for plagiarism. In 2017, Charles Armstrong, once a leading figure in Korean scholarship, returned the 2014 John King FairbankPrize from the American Historical Society after allegations emerged that he had plagiziared widely in his book,“Tyranny of the Weak: North Korea and the World, 1950–1992.”At the time, Armstrong admitted to having made “citation errors” in the work. However, Balazs Szalontai, anacademic in Korea, insisted that the the errors were in fact plagiarism and that they were sweeping. Now, in what Szalotai told us was the earliest instance of Armstrong’s plagiarism that he has found, the journal *ColdWar History* is retracting an article by Armstrong. According to the notice:*We, the Editors and Publisher of* Cold War History*, have retracted the following article:Charles K. Armstrong, ‘“Fraternal Socialism”: The International Reconstruction of North Korea,1953-62’,* Cold War History*, 5, 2 (2005) 161-187*

Fourth retraction for Haruko Obokata, focus of STAP cell scandal, after Harvard investigation. More than five years after *Nature* retracted two highly suspect papers about what had been described as a major breakthrough in stem cell research, another journal has pulled a paper about the work.The scandal over so-called STAP stem cells took down more than just a few articles. The case centered on Haruko Obokata, a Japanese researcher who conducted the studies as a post-doc in the Harvard lab of Charles Vacanti. Obokata lost her doctoral thesis from Waseda University in 2015 because it plagiarized from the U.S. National Institutes of Health. She also retracted a paper in *Nature Protocols*. Vacanti, who had been one of the most prominent figures in regenerative medicine, ended up leaving Harvard, where he had been the chair of the Department of Anesthesiology, Perioperative and Pain Medicine. He appears tohave dropped out of science entirely.Tragically, one of Obokata’s co-authors and a former mentor, Yoshiki Sasai, committed suicide as a result of the revelations about the findings emerged. The latest article to fall, titled “The potential of stem cells in adult tissues representative of the three germ layers,”appeared in Tissue Engineering Part A. According to the notice:*Dr. Charles A. Vacanti, the senior author of the article entitled, “The Potential of Stem Cells in Adult Tissues Representative of the Three Germ Layers,” (vol. 17, nos. 5/6, pages 607–615),published in the March 2011 issue of* Tissue Engineering*:* Part A *contacted the Journal to request a formal retraction of the article following an investigation by a Harvard Medical School Committee on Scientific Integrity evaluating a published erratum of the original work presented.\ The reader is directed to the published correction notice to the article in 2014. While the* *authors maintained that the published erratum was accurate, the committee could not validate the accuracy of the corrections. Consequently, the authors agreed to request a formal retraction.*

Famous retraction – STAP Cell   Read GOODYEAR

Acid bath stem-cell scientist can't reproduce results

15:34 19 December 2014 by Helen Thomson
For similar stories, visit the Stem Cells Topic Guide

"I can't find the words to apologise," says Haruko Obokata, at the Riken Institute in Kobe, Japan, who today admitted failing in an attempt to reproduce her infamous "STAP" cell experiments.

The buck STAPs here *(Image: Kyodo/Reuters)*

An eight month investigation by a team of Riken researchers, performed under the guidance of Obokata, attempted to reproduce experiments first published in two papers in January. The papers claimed that almost any adult cell could be coaxed into becoming a stem cell just by dipping it in a bath of acid for 30 minutes. The method held great promise for regenerative medicine because it would make it possible to create any cell without reprogramming genes, or destroying an embryo. Obokata's team called this technique stimulus–triggered acquisition of pluripotency, or STAP.

https://www.newyorker.com/magazine/2016/02/29/the-stem-cell-scandal

RETRACTIONS Shaking Up Science : Science 25 January 2013: Vol. 339 no. 6118 pp. 386-389   Jennifer Couzin-Frankel: Fang, Steen, and Casadevall "assembled an enormous Excel file of every retraction they could find in PubMed, more than 2000 dating back to 1977. They cross-referenced many retractions with other sources, such as reports from the U.S. Office of Research Integrity (ORI), which investigates misconduct. **They attributed about 67% of all the retractions to scientific misconduct, including fraud and plagiarism.** The results were published in October 2012 PNAS. "We never anticipated that the problem was going to be so widespread, ever," says Casadevall, who'd expected honest errors to explain the vast majority of retractions. "We need to clean up our act." ORI reports show gender imbalance: only 9 of 72 faculty members were women –1/3d number one-third based on female % life sci.   This seems to contradict Steen's report in Journal of Medical Ethics that 73.5% of 742 papers retracted between 2000 and 2010 were pulled because of errors. Why? Steen based analysis on reasons AUTHORS gave for retraction, rather than tracking down story.

Bik EM, Casadevall A, Fang FC. 2016. The prevalence of inappropriate image duplication in biomedical research publications. mBio 7(3):e00809-16. doi:10.1128/mBio.00809-16. This study attempted to determine the percentage of published papers that contain inappropriate image duplication, a specific type of inaccurate data. The images from a total of 20,621 papers published in 40 scientific journals from 1995 to 2014 were visually screened. Overall, 3.8% of published papers contained problematic figures, with at least half exhibiting features suggestive of deliberate manipulation. **The prevalence of papers with problematic images has risen markedly during the past decade. Additional papers written by authors of papers with problematic images had an increased likelihood of containing problematic images as well.** As this analysis focused only on one type of data, it is likely that the actual prevalence of inaccurate data in the published literature is higher. The marked variation in the frequency of problematic images among journals suggests that journal practices, such as pre publication image screening, influence the quality of the scientific literature.

Science ranked first for articles retracted, 70, just edging out PNAS, which comes second with 69. Thirty-two of Science's retractions were due to fraud or suspected fraud, and 37 to error. Why? Incentive to publish in those journals leads to errors/fraud. Journals want headline paper. Worth checking paper. Casadevall and Fang devised a "retraction index" to show that journals with relatively high impact factors, such as Science, Nature, and Cell, had a higher rate of retractions.   But Although retractions are on the rise, they remain relatively rare in science. Well under 0.1% of papers in PubMed have been retracted.

Economics treats all of these decisions as responses to incentives, in which for enough reward many/most/ almost all people would put aside ethics and go down the slippery slope. As the million $ man says "Everyone has his price". In any tournament or situation where cheating can gain you higher income/status, it is rational to compare the benefit from cheating with the chance of getting caught and penalties:

Commit crime/falsify results if (1-p)W(falsified paper) > p C where p is chance of getting caught, W is gain from falsification, C is Cost of falsification. If p=0 or C=0 , you falsify. Likely that p goes up with importance of paper – stem cells from being in acid brine → efforts to use/improve/xx result. If paper is obscure copy of obscure, much lower p. Does this mean papers with 0 cites in low impact journal more likely to be falsified/plagiarized?

More On Retractions Furman, Jensen and Murray Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine Research Policy Volume 41, Issue 2, March 2012, Pages 276-290 –

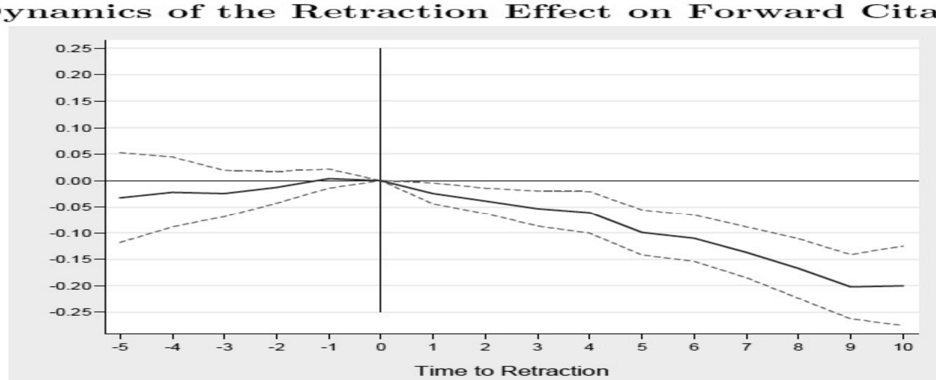
- Study Design
  - Articles retracted due to allegations of falsification
    - Jan 1 1980 - March 1 2006
  - Data from journals with greater than 10 and greater than 30,000 annual citations
- Results
  - 14 eligible journals had 63 eligible retracted articles
  - Median time from publication to retraction was 28 months
    - 79 months for articles from senior researchers
    - 22 months for junior researchers
  - Retracted articles and controls had the same number of citations for the first 12 months
  - Retracted papers were twice as likely to have multinational authorship

Pfeifer, Snodrass The Continued Use of Retracted, Invalid Scientific Literature -JAMA,1990;263(10):1420-1423.

Summary
- Identified 82 retracted papers
- Examined their citation after retraction
- 3 of the papers were cited 733 times
- Comparison with a control group found that retraction reduced citation by 35%
- Although after retractions US authors are less likely to sight a paper they are the largest originating source for retractions
  - Why not 100% fall in citations
    - Information from retracted papers is still available
    - Different journals have different formats for retraction
    - Lack of attention to manuscripts by authors and editors

Impact of false science on Science: RETRACTIONS Azoulay, Furman, Krieger,Murray NBER WP 18499 2014 revision examine the impact of more than 1,100 scientific retractions on the citation trajectories of articles that are **related to retracted papers in intellectual space but published prior to the retraction event**. Following retraction and relative to … controls, related articles experience a lasting 5-10% decline in rate of citations received.

Figure 3: Dynamics of the Retraction Effect on Forward Citation Rates



Note: The solid blue lines in the above plot correspond to coefficient estimates stemming from conditional fixed effects quasi-maximum likelihood Poisson specifications in which the citation rates for related articles and their controls are regressed onto year effects, article age indicator variables, as well as 20 interaction terms between treatment status and the number of years before/elapsed since the retraction event (the indicator variable for treatment status interacted with the year of retraction itself is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around case codes) around these estimates is plotted with dashed red lines.

In addition, **the arrival rate of new articles and funding** into these fields decrease after a retraction. The evidence is consistent with the view that scientists avoid retraction-afflicted fields lest their reputation suffer through mere association, but cannot rule out the possibility that our estimates also reflect scientists' learning about the fields' shaky intellectual foundations.

- Results
  - Commercial citers are less responsive to the retraction event than academic citers
  - Penalty suffered by related articles is greater when the associated retraction includes fraud or misconduct, relative to honest mistakes.

Table 9: Interpreting Citation Behavior for Articled Related to "Absent Shoulders" Retractions

	Retracted Papers				Related Papers			
	(1a)	(1b)	(1c)	(1d)	(2a)	(2b)	(2c)	(2d)
	Academic Citations Only	Private-Firms Citations Only	Academic Citations Only	Private-Firms Citations Only	Academic Citations Only	Private-Firms Citations Only	Academic Citations Only	Private-Firms Citations Only
After Retraction	-1.293**	-1.309**	-1.304**	-1.283**	-0.054**	-0.006	-0.071**	-0.005
	(0.154)	(0.188)	(0.180)	(0.236)	(0.017)	(0.023)	(0.017)	(0.025)
After Retraction × Retracted Paper Cited in Patent			0.041	-0.086				
			(0.178)	(0.328)				
After Retraction × Related Paper Cited in Patent							0.066†	-0.000
							(0.038)	(0.045)
Nb. of Retraction Cases	304	304	304	304	334	334	334	334
Nb. of Source Articles	1,089	1,089	1,089	1,089	589	589	82,819	53,357
Nb. of Related/Control Articles					62,205	62,205	96,373	61,806
Nb. of Article-Year Obs.	15,711	15,711	15,711	15,711	807,203	807,203	1,238,118	801,709
Log Likelihood	-30,568	-8,234	-30,568	-8,234	-1,366,136	-402,337	-1,756,286	-400,178

Note: Estimates stem from conditional quasi-maximum likelihood Poisson specifications. The dependent variable is the total number of forward citations (exclusive of self-citations) received by each related article in a particular year. All models incorporate a full suite of year effects as well as 31 article age indicator variables (age zero is the omitted category). Exponentiating the coefficients and differencing from one yields numbers interpretable as elasticities.

In columns (2a) and (2b), the estimation sample is limited to those related articles and controls that receive at least one "private firm" citation between their year of publication and 2011. For this analysis, a citation is said to emanate from a private firm when at least one address listed by the *Web of Science* includes a suffix such as Inc., Corp., LLC, Ltd., GmbH, etc.

QML (robust) standard errors in parentheses, clustered around retraction cases.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Attention is a key predictor of retraction – retracted articles arise most frequently among highly-cited articles. Finds false stuff fast, not systematically affected by author prominence, and produce immediate, severe, and long-lived decline in future citations. Most robust and statistically significant predictors of retraction are (a) the association of the corresponding author with a top US research university and (b) the number of citations that an article receives in its first year after publication. So system works well ??---Furman, Jensen, Murray

Jones and Uzzi: Retracted papers citations fall and so too past papers of people, with differences between those where self-reported and other person causing the retraction

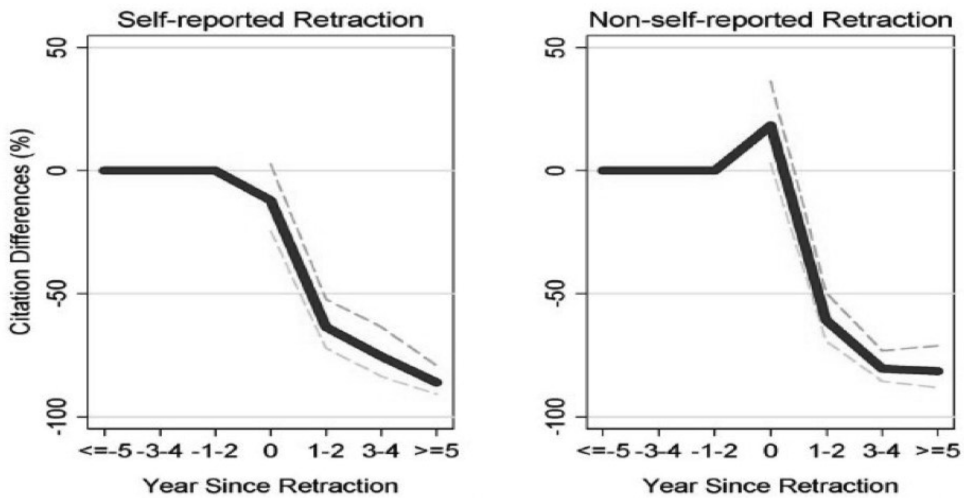
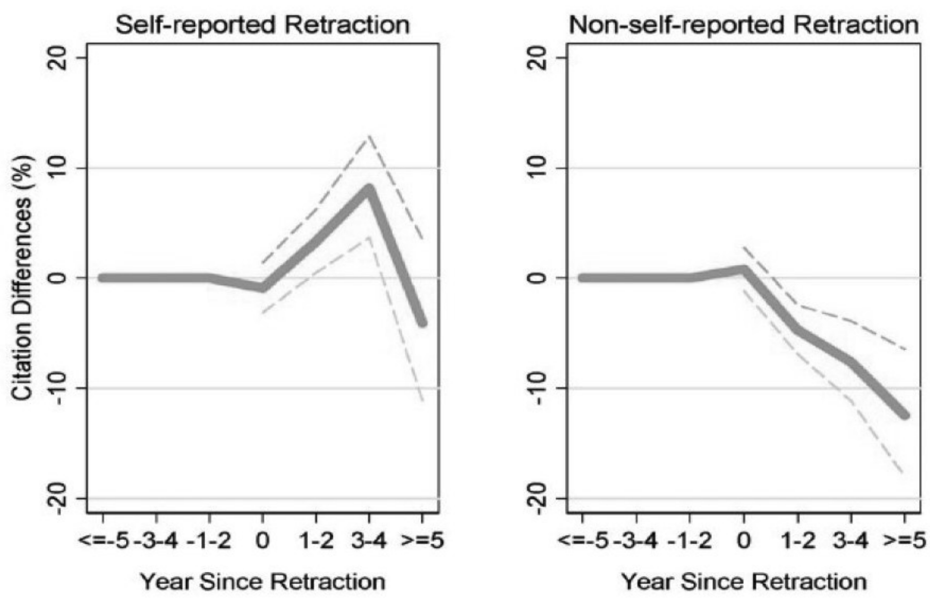


Figure 2 | Effect of retraction on retracted papers themselves. Citations losses, compared to control papers, are shown for (a) self-reported retra and (b) non-self-reported retractions. Blue lines indicate mean citation losses and dashed lines present 95% confidence intervals. Compared to control papers, citation losses are 86.2% ($p < .0001$) for self-reported retractions and 81.5% ($p < .0001$) for non-self-reported retractions, annual or more years after retraction.



igure 3 | **Effect of retraction on authors' prior body of work.** Citations losses for prior work, compared to control papers, are presented after (a) self-eported retractions and (b) non-self-reported retractions. Orange lines indicate mean citation losses and dashed lines present 95% confidence ntervals. After non-self-reported retractions, the authors' prior work loses 12.5% ($p < .0001$) of citations per year per prior publication five or more years fter the retraction event, compared to control papers. By contrast, citation losses for the authors' prior body of work do not appear after self-reported etractions.

Azoulay, et al 2015 Relative to non-retracted control authors, faculty members who experience a retraction see the citation rate to their previous work drop by 10% on average ... We then investigate whether the eminence of the retracted author, and the publicity surrounding the retraction, shape the magnitude of the penalty. … eminent scientists are more harshly penalized than their less-distinguished peers in the wake of a retraction, but only in cases involving fraud or misconduct. When the retraction event had it source in "honest mistakes," we find no evidence of differential stigma between high- and low-status faculty members.

And no retraction

Wrong Data: March 2, 2016 The U.S. Dietary Guidelines: A Scientific Fraud

Archer "The Scientific Report of the 2015 Dietary Guidelines Advisory Committee was primarily informed by memory-based dietary assessment methods (M-BMs) (eg, interviews and surveys). The reliance on M-BMs to inform dietary policy continues despite decades of unequivocal evidence that M-BM data bear little relation to actual energy and nutrient consumption."

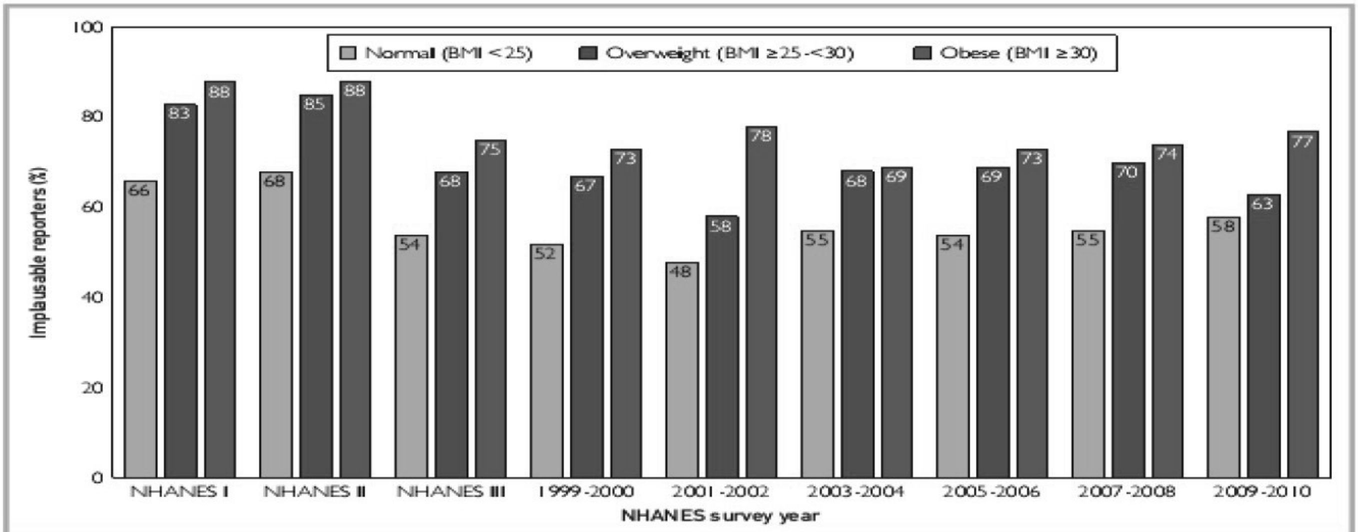
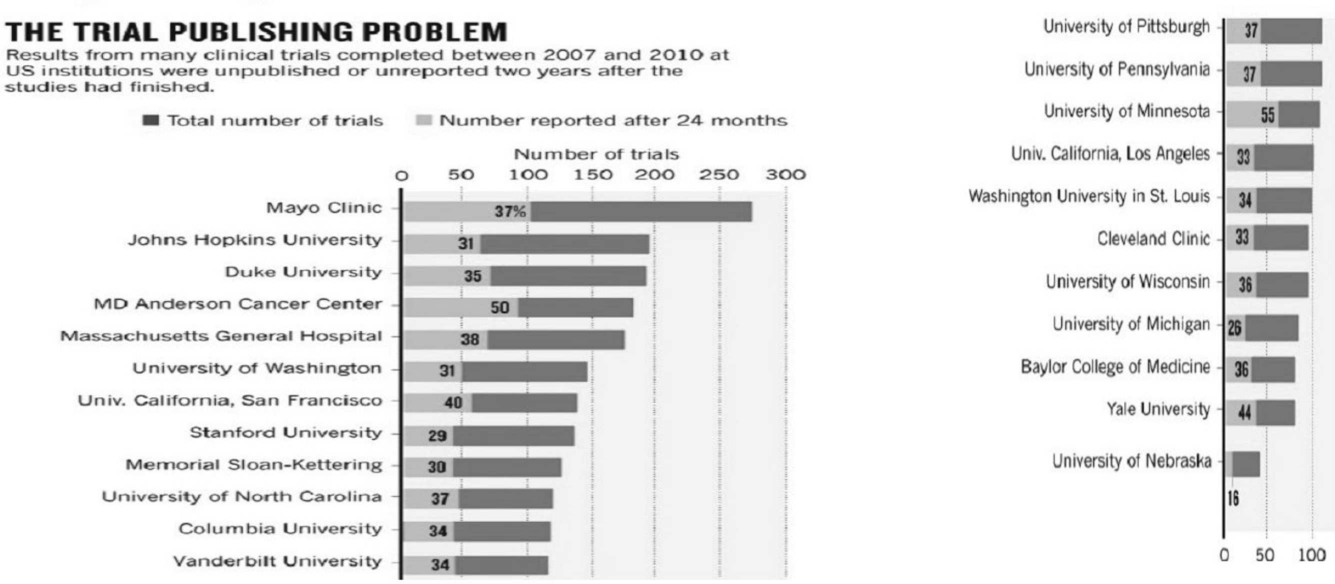


FIGURE. Percentage of implausible reporters by body mass index (BMI) for US women aged 20 to 74 years in the National Health and Nutrition Examination Survey (NHANES) (1971-2010). Physiologically implausible values were determined via the following equation: (reported energy intake/basal metabolic rate) ≤1.35. Implausible values may be considered "incompatible with life."[11,p7]

Shoddy science

Study Finds SLIGHTLY MORE THAN HALF OF PAPERS SUPPORT THE IDEA THAT SALT IS BAD FOR YOUR HEALTH Analysis of scientific reports and comments on the health effects of a salty diet reveals a polarization between those supportive of the hypothesis that population-wide reduction of salt intake is associated with better health and those that were not. In all, 54 percent were supportive of the hypothesis; 33 percent, not supportive; and 13 percent inconclusive. A citation analysis found papers on either side of the hypothesis were more likely to cite reports that drew a similar conclusion than to cite reports drawing a different conclusion. Dominating the literature were a small number of influential papers that presented strong evidence for and against. Fourth, there was very little consist-ency in the selection of primary studies in systematic re-views, further compounding the challenges the field faces in achieving resolution. The literature contains two almost distinct and disparate lines of scholarship, one supporting and one contradicting the hypothesis that salt reduction will improve clinical outcomes.

Failure to report in timely fashion



- Results
    - Commercial citers are less responsive to the retraction event than academic citers
    - Penalty suffered by related articles is greater when the associated retraction includes fraud or misconduct, relative to honest mistakes.

Table 9: Interpreting Citation Behavior for Articled Related to "Absent Shoulders" Retractions

	Retracted Papers				Related Papers			
	(1a)	(1b)	(1c)	(1d)	(2a)	(2b)	(2c)	(2d)
	Academic Citations Only	Private-Firms Citations Only	Academic Citations Only	Private-Firms Citations Only	Academic Citations Only	Private-Firms Citations Only	Academic Citations Only	Private-Firms Citations Only
After Retraction	-1.293**	-1.309**	-1.304**	-1.283**	-0.054**	-0.006	-0.071**	-0.005
	(0.154)	(0.188)	(0.180)	(0.236)	(0.017)	(0.023)	(0.017)	(0.025)
After Retraction × Retracted Paper Cited in Patent			0.041	-0.086				
			(0.178)	(0.328)				
After Retraction × Related Paper Cited in Patent							0.066†	-0.000
							(0.038)	(0.045)
Nb. of Retraction Cases	304	304	304	304	334	334	334	334
Nb. of Source Articles	1,089	1,089	1,089	1,089	589	589	82,819	53,357
Nb. of Related/Control Articles					62,205	62,205	96,373	61,806
Nb. of Article-Year Obs.	15,711	15,711	15,711	15,711	807,203	807,203	1,238,118	801,709
Log Likelihood	-30,568	-8,234	-30,568	-8,234	-1,366,136	-402,337	-1,756,286	-400,178

Note: Estimates stem from conditional quasi-maximum likelihood Poisson specifications. The dependent variable is the total number of forward citations (exclusive of self-citations) received by each related article in a particular year. All models incorporate a full suite of year effects as well as 31 article age indicator variables (age zero is the omitted category). Exponentiating the coefficients and differencing from one yields numbers interpretable as elasticities.

In columns (2a) and (2b), the estimation sample is limited to those related articles and controls that receive at least one "private firm" citation between their year of publication and 2011. For this analysis, a citation is said to emanate from a private firm when at least one address listed by the *Web of Science* includes a suffix such as Inc., Corp., LLC, Ltd., GmbH, etc.

QML (robust) standard errors in parentheses, clustered around retraction cases.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

REPLICATION CRISIS IN SEVERAL FIELDS

Inability to replicate --John P. A. Ioannidis PLoS Med. 2005 Aug; 2(8): e124. **Why Most Published Research Findings Are False** … Answer because focus on isolated discoveries by single teams and interpret research experiments in isolation. One study claims a research finding, and this receives unilateral attention. The probability that at least one study, among several done on the same question, claims a statistically significant research finding is easy to estimate.

Let R be the ratio of the number of "true relationships" to "no relationships"The probability of finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists is Type I error rate, α. Assuming that c relationships are being probed expected values are:

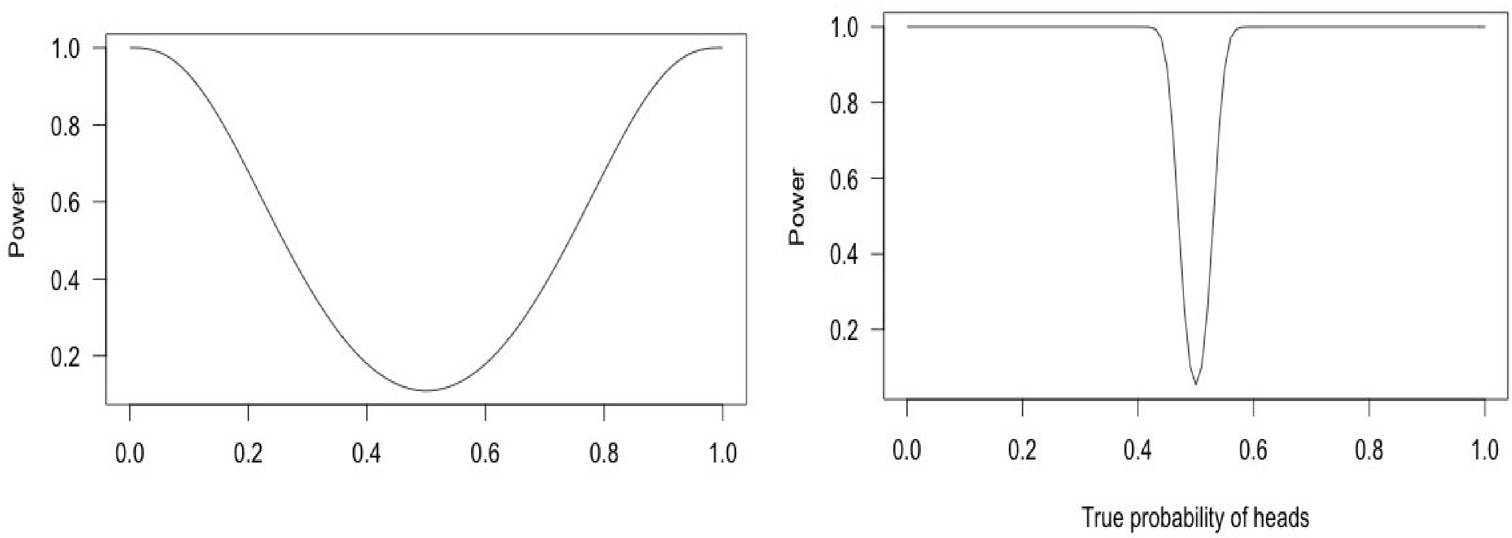
Table 1. Research Findings and True Relationships

Research Finding	True Relationship: Yes	No	Total
Yes	$c(1-\beta)R/(R+1)$	$c\alpha/(R+1)$	$c(R+\alpha-\beta R)/(R+1)$
No	$c\beta R/(R+1)$	$c(1-\alpha)/(R+1)$	$c(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

DOI: 10.1371/journal.pmed.0020124.t001

Problem of Under-powered Studies – too few observations to make an assessment. Example: people have no preference between A and B. But if you give them **A** and b, they choose **A.** With sample of 10 you have little chance that **A** in **Ab** will be chosen more than A in B if preference is modest. With sample of 1000 you can detect small differences well. But purely by chance can get **A** in **Ab** will chosen more than A in B even if your treatment has no effect. People can report exciting under-powered results – all of us do experiment with 10 and those of us getting no effects or negative effects, say nothing. But the person who gets 8 of 10 yells hurray.

Chance of Deciding Effect different than random with 10 vs 1000



In one sample of studies published between 1975 and 1990 in prestigious medical journals, 27% of randomized controlled trials gave negative results, but 64% of these didn't collect enough data to detect a 50% difference in primary outcome between treatment groups. Even if one medication decreases symptoms by 50% more than the other medication, there's insufficient data to conclude it's more effective. And 84% of the negative trials didn't have the power to detect a 25% difference. That's not to say scientists are lying when they state they detected no significant difference between groups. You're misleading yourself when you assume this means there is no real difference. There may be a difference, but the study was too small to notice it. https://www.statisticsdonewrong.com/power.html

Brodeur et al Using 50,000 tests published in the AER, JPE, and QJE, we identify a residual in the distribution of tests that cannot be explained solely by journals favoring rejection of the null hypothesis. We observe a two humped camel shape with missing p values between 0.25 and 0.10 that can be retrieved just after the 0.05 threshold and represent 1020 percent of marginally rejected tests. Our interpretation is that researchers inflate the value of just rejected tests by choosing "significant" specifications. We propose a method to measure this residual and describe how it varies by article and author characteristics.

The accuracy of published medical research is critical both for scientists, physicians and patients who rely on these results. But the fundamental belief in the medical literature was called into serious question by a paper suggesting most published medical research is false. Here we adapt estimation methods from the genomics community to the problem of estimating the rate of false positives in the medical literature using reported P-values as the data. We then collect P-values from the abstracts of all 77,430 papers published in *The Lancet, The Journal of the American Medical Association, The New England Journal of Medicine, The British Medical Journal,* and *The American Journal of Epidemiology* between 2000 and 2010. We estimate that the overall rate of false positives among reported results is 14% (s.d. 1%), contrary to previous claims. We also find there is not a significant increase in the estimated rate of reported false positive results over time (0.5% more FP per year, P = 0.18) or with respect to journal submissions (0.1% more FP per 100 submissions, P = 0.48). Statistical analysis must allow for false positives in order to make claims on the basis of noisy data. But our analysis suggests that the medical literature remains a reliable record of scientific progress.

Reproducibility and incentives to correct

NATURE COMMENT Reproducibility : A tragedy of errors 03 February 2016 Mistakes in peer-reviewed papers are easy to find but hard to fix, report David B. Allison, Andrew W. Brown, Brandon J. George & Kathryn A . Kaiser

The authors say that, in the course of assembling weekly lists of articles on obesity and nutrition, they began to notice more peer-reviewed articles containing what they refer to as 'substantial or invalidating errors.' "What was striking was how severe some of these errors were, involving mathematically impossible values, probabilities greater than one, weight loss results that, if true, would have required that adults had grown over 6 centimeters in height in two months, to name just a few," "These errors involved factual mistakes or practices which veered substantially from clearly accepted procedures in ways that, if corrected, might alter a paper's conclusions. In several cases, our noting these errors led to retractions of the papers containing them." Brown says the team attempted to address more than 25 of these errors with letters to authors or journals. Their efforts revealed invalidating practices that occur repeatedly and showed how journals and authors react when faced with mistakes that need correction.Post-publication peer review is not consistent, smooth or rapid. Many journal editors and staff seemed unprepared to investigate, take action or even respond. Too often, the process spiraled through layers of ineffective emails among authors, editors and unidentified journal representatives, often without any public statement's being added to the original article." During the informal 18-month review of literature, the authors found a number of recurring problems:

- Editors are often unprepared or reluctant to take speedy and appropriate action
- Where to send expressions of concern is unclear
- Journal staff who acknowledged invalidating errors were reluctant to issue retractions or timely expressions of concern
- Some journals may charge fees to authors who report the issues to correct others' mistakes (more than $1,000)
- No standard mechanism exists to request raw data for review to confirm the errors
- Concerns expressed through online forums are easily overlooked.

Penalties

NATURE | NEWS

US vaccine researcher sentenced to prison for fraud

The case of Dong-Pyou Han illustrates the uneven nature of penalties for scientific misconduct.

Soon to come: machine learning to detect possible fraudulent work

Markowitz and Hancock examine whether scientists write differently when reporting on fraudulent research. In an analysis of over two million words, we evaluated 253 publications retracted for fraudulent data and compared the linguistic style of each paper to a corpus of 253 unretracted publications and 62 publications retracted for reasons other than fraud (e.g., ethics violations). Fraudulent papers were written with significantly higher levels of linguistic obfuscation, including lower readability and higher rates of jargon than unretracted and nonfraudulent papers. We also observed a positive association between obfuscation and the number of references per paper, suggesting that fraudulent authors obfuscate their reports by making them more costly to analyze and evaluate.

Earlier paper "Linguistic Traces of Scientific Fraud" looked at social psychologist Diederik Stapel and found "more terms pertaining to methods, investigation, and certainty" than in his genuine papers

Reconstruction of a Train Wreck: How Priming Research Went off the Rails

February 2, 2017 **Kahneman**, Priming, r-index, Statistical Power, Thinking Fast and Slow

Authors: Ulrich Schimmack, Moritz Heene, and Kamini Kesavan

Abstract:
We computed the R-Index for studies cited in Chapter 4 of Kahneman's book "Thinking Fast and Slow." This chapter focuses on priming studies, starting with John Bargh's study that led to Kahneman's open email. The results are eye-opening and jaw-dropping. The chapter cites 12 articles and 11 of the 12 articles have an R-Index below 50. The combined analysis of 31 studies reported in the 12 articles shows 100% significant results with average (median) observed power of 57% and an inflation rate of 43%. The R-Index is 14. This result confirms Kahneman's prediction that priming research is a train wreck and readers of his book "Thinking Fast and Slow" should not consider the presented studies as scientific evidence that subtle cues in their environment can have strong effects on their behavior outside their awareness.

According to the Schimmack et al blog,

> …readers of his [Kahneman's] book "Thinking Fast and Slow" should not consider the presented studies as scientific evidence that subtle cues in their environment can have strong effects on their behavior outside their awareness.

Remarkably, Kahneman took the time to post a detailed response to the blog, writing:

> What the blog gets absolutely right is that I placed too much faith in underpowered studies. As pointed out in the blog, and earlier by Andrew Gelman, there is a special irony in my mistake because the first paper that Amos Tversky and I published was about the belief in the "law of small numbers," which allows researchers to trust the results of underpowered studies with unreasonably small samples. We also cited Overall (1969) for showing "that the prevalence of studies deficient in statistical power is not only wasteful but actually pernicious: it results in a large proportion of invalid rejections of the null hypothesis among published results." Our article was written in 1969 and published in 1971, but I failed to internalize its message.

We contacted Kahneman, who confirmed that he indeed posted this response on the blog Replicability-Index.

What is the R index? A statistical measure designed to measure likelihood an exact replica will give significant result found in earlier study https://replicationindex.com/

Dr. Ulrich Schimmack's Blog about Replicability

Sticky Bayes-Factor, Default-Baysian-t-test, Magic, Median Observed Power, Meta-Analysis, Observed Power, Pcurve, Post-Hoc Power, Posteriori Power Analysis, Power, Psychology, Puniform, r-index, R-Index science R-Index4Science, Replicability, Replicability Ranking, Replicability Report, Yuan and Maxwell Magic, Magic-Index

"For generalization, psychologists must finally rely, as has been done in all the older sciences, on replication" **(Cohen, 1994).**

***DEFINITION OF REPLICABILITY:** In empirical studies with sampling error, replicability refers to the probability of a study with a significant result to produce a significant result again in an exact replication study of the first study using the same sample size and significance criterion (Schimmack, 2017).*