

Ten Simple Rules for Scientific Fraud & Misconduct

Nicolas P. Rougier (Inria) & John Timmer (Ars Technica)

March 7, 2019

Disclaimer

I obviously do not encourage scientific fraud nor misconduct. The goal of this talk is to alert the public to problems that have arisen in part due to the Publish or Perish imperative, which has driven a number of researchers to cross the Rubicon without the full appreciation of the consequences. Choosing fraud will hurt science, end careers, and could have impacts on life outside of the lab. If you're tempted (even slightly) to beautify your results, keep in mind that the benefits are probably not worth the risks.

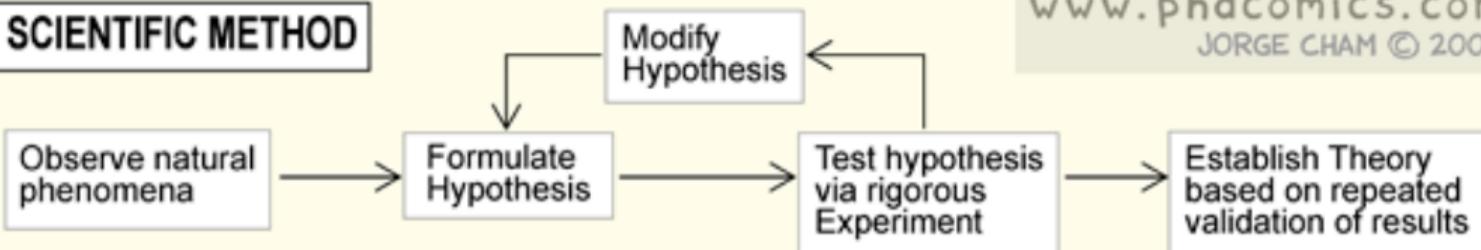


In the news (2018)

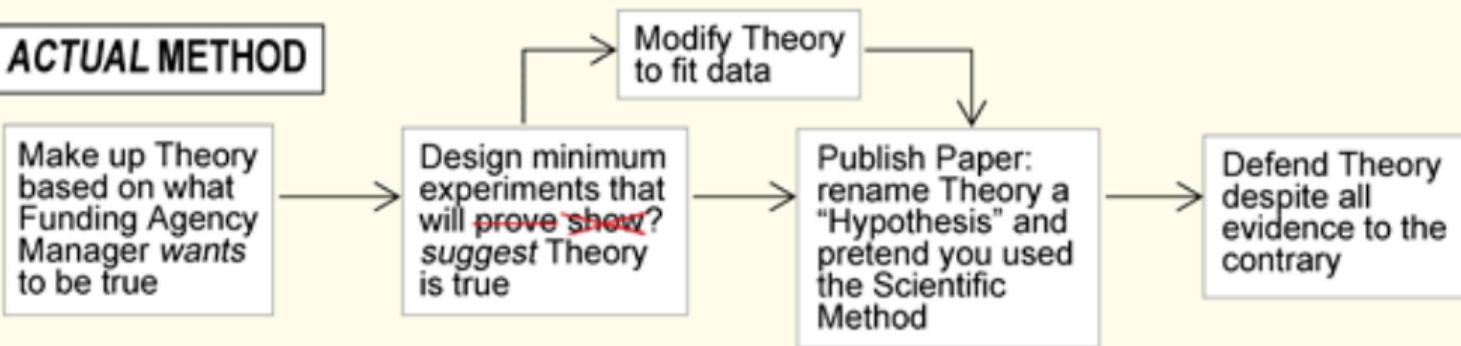
- Nope, I won't share my data (NASA)
Two years of stonewalling, Nathan Myhrvold, June 2018
- The Stanford prison experiment was a fraud
The Lifespan of a Lie, Ben Blum, June 2018
- A 100M\$, 10 years study canceled
It Was Supposed to Be an Unbiased Study of Drinking, Roni Caryn Rabin, June 2018
- Does the Mediterranean diet prevent heart attacks? (no)
Why a major paper on the Mediterranean diet was just retracted and replaced, Samantha Schmidt, June 2018



THE SCIENTIFIC METHOD



THE ACTUAL METHOD



Rule 1: Misrepresent, falsify, or fabricate your data

In order to start your life as a scientific fraudster, the first thing you need to do is learn how to convincingly misrepresent, falsify, or fabricate data.



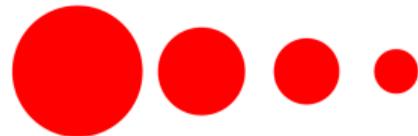
Relative size using full range

Relative size using partial range



Relative size using disc area

Relative size using disc radius



Ten Simple Rules for Better Figures

Famous data makers & transformers

Andrew Wakefield, The Lancet, 1998

→ *based not on bad science but on a deliberate fraud*

Jan Hendrik Schön, Nature, Science, 2000

→ *forged results by using the same data sets for different and unrelated experiments*

Olivier Voinnet, Plant Cell, 2015

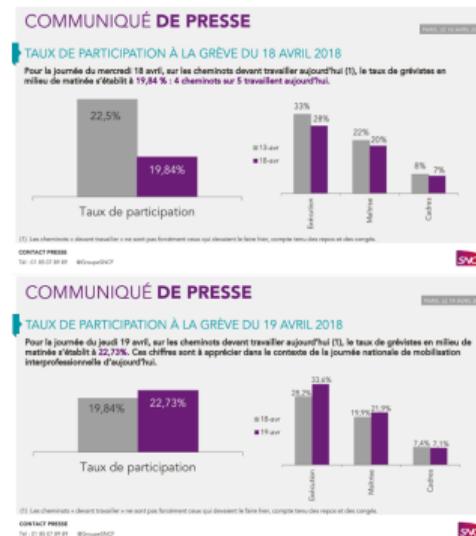
→ *multiple cases of data manipulation*

Michael Lacour, Science, 2015

→ *numerous irregularities in the methodology and falsified data*

Oona M. Lönnstedt & Peter Eklöv, Science, 2016

→ *absence of original data for the experiments reported in the paper*



Everything's possible with SNCF

... nous observons d'ores et déjà, en termes de pré-inscription en licence, une augmentation de 11% pour les étudiants du Sénégal, de 5,6% pour les étudiants du Mali, et encore de 8% pour les étudiants du Bénin. Globalement il y a une sorte de stabilité dans ces pré-inscriptions, ce qui fait qu'une fois de plus les craintes s'avèrent infondées.

— Frédérique Vidal, Ministry of Higher Education, Research and Innovation, 2019

Au global, la comparaison entre les chiffres provisoires 2019 et les chiffres de fin de procédure 2018 met en lumière une baisse de -10% des candidats ayant soumis leur dossier...

— Campus France, 2019

Rule 2: Hack your results

If you are reluctant to manipulate your data, you still have the option of searching through your results to find anything that reaches significance (a.k.a. p-hacking).

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥0.1	

XKCD #1478

Enter ten random numbers between 01 and 100

1	2
<input type="text" value="82"/>	<input type="text" value="58"/>
3	4
<input type="text" value="7"/>	<input type="text" value="34"/>
5	6
<input type="text" value="28"/>	<input type="text" value="66"/>
7	8
<input type="text" value="80"/>	<input type="text" value="19"/>
9	10
<input type="text" value="28"/>	<input type="text" value="43"/>

Unbelievable!

You chose the numbers 82 58 07 34 28 66 80 19 28 43

But these are clearly not random numbers. We can tell because random numbers do not contain patterns but the numbers you entered show a fairly obvious pattern.

Take another look at the sequence you put in. You will see that the number of times that '8' appears in the sequence is: 5. But the 'expected number' from a random process is just 2. How odd is this pattern? Quite odd in fact. The probability that a truly random process would turn up numbers like this is just $p=0.0427$ (i.e. less than 5%).

Try again (with really random numbers this time)!

ps: you might think that if the p value calculated above is high (for example if it is greater than 12%) that this means that the numbers you chose are not all that odd; but in fact it means that the numbers are really particularly odd since the probability that the fishy test would produce a p values above 12%, when really random sequences are used, is low ($p<0.07$). For more on how to fish see [here](#).

From <http://egap.org/content/go-fishing>

Cornell's #pizzagate: p-hacking + HARKing

The rise and fall of a food behavior scientist:

When she arrived, I gave her a data set of a self-funded, failed study which had null results (it was a one month study in an all-you-can-eat Italian restaurant buffet where we had charged some people ½ as much as others). I said, "This cost us a lot of time and our own money to collect. There's got to be something here we can salvage because it's a cool (rich & unique) data set..."

—Brian Wansink, *The Grad Student Who Never Said "No"*, 2016

- How Traumatic Violence Permanently Changes Shopping Behavior [Retracted]
- How Pricing Influences Regret at All-You-Can-Eat Buffets [Retracted]
- Eating Heavily: Men Eat More in the Company of Women
- Peak-end Pizza: Prices Delay Evaluations of Quality
- Lower Buffet Prices Lead to Less Taste Satisfaction [Retracted]

Rule 3: Copy/paste from others

It's not enough to have (fake) results; you still need to publish them if you want to gain fame But why bother writing? It's much simpler to copy/paste what others have written.

“On dit que Watson et torticolis ont découvert l'ADN alors qu'en fait c'est Rosalind Franklin”



Types of plagiarism

- Resignation in 2011 of Karl-Theodor zu Guttenberg as Minister of Defence of Germany and revocation of his doctorate over the plagiarism of his doctoral dissertation
- Gilles Bernheim resigned as chief rabbi in 2013 amid revelations of plagiarism and deception about his academic credentials
- Etienne Klein fired from the head of the IHEST (2017) over allegation of plagiarism

→ <http://archeologie-copier-coller.com>

A Natural History of Skills 1

A Natural History of Skills

Thomas Bouraud^{1,2}, Arthur Leblais³, Nicolas P. Rougier^{4,5}

¹CNRS, UMR 5293, IMN, 33000 Bordeaux, France
²University of Bordeaux, UMR 5293, IMN, 33000, Bordeaux, France
³CNRS, French-Israeli Neuroscience Lab, 33000 Bordeaux, France
⁴CHU de Bordeaux, IMN Clinique, 33000 Bordeaux, France
⁵IMIA Bordeaux Sud-Ouest, 33405, Talence, France

⁶IPB, University of Bordeaux, IPB, CNRS, UMR 5800, 33405 Talence, France

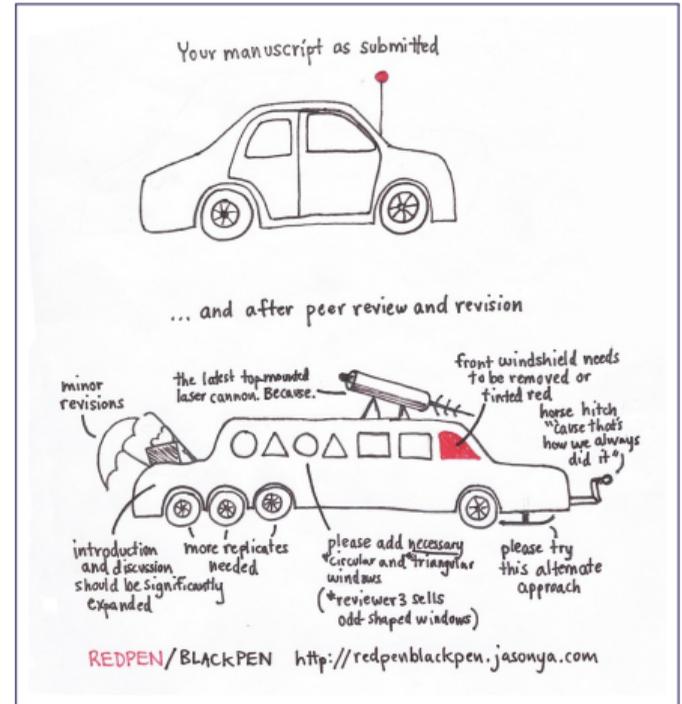
⁷These authors contributed equally to this work.

Abstract	2
1. Introduction	3
1.1. The function(s) of the forebrain	4
1.2. Habits, automaticity & skills	5
2. Multiple systems at work	6
2.1. The mammalian cortex	7
2.2. The thalamus and the basal ganglia	8
2.3. Skill learning and synaptic plasticity	10
2.3.1 Hebbian learning	11
2.3.2. Dopamine and reinforcement learning	11
3. Parallel emergence across evolution	11
3.1. Lamprey	12
3.2. Fishes, frogs & reptiles	14
3.3. Mammals	15
3.4. Birds	16
Box 1	17
4. The automatization of skills	19
4.1. Distinct but complementary roles	19
4.2. A counter-intuitive hypothesis	20
4.3. Compatibility with the alternative hypotheses	20
4.4. The entanglement of two systems	22
4.5. Computational Approach	22
6. Conclusion	25
Acknowledgement	26
Bibliography	27

No consequence but annoying

Rule 4: Write your own peer-review

If reviewers are picky, they may ask annoying questions, request more information or experiments, ask for a major revision and/or reject your submission entirely. Fortunately, there is a simple solution: write your own review!



Can you spot the fake review?

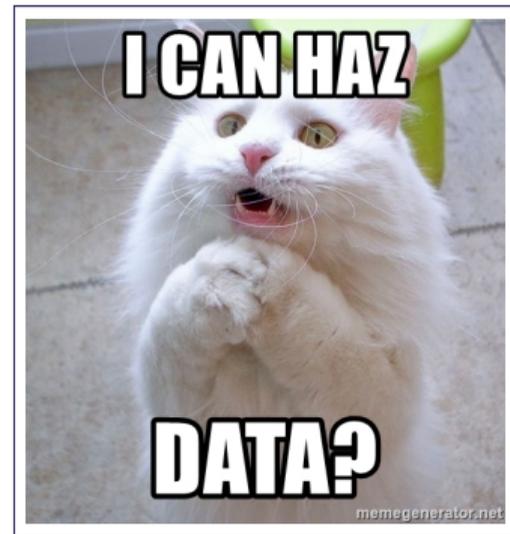
- Sorry for our long silence, due to some perplexity on our side at reading your manuscript
- The paper is presented as a rather undigestible and tortuous collection of disparate results
- In order to be able to publish this manuscript it needs to be rewritten in the form of a scientific article
- The biggest problem with this manuscript, which has nearly sucked the will to live out of me, is the terrible writing style.
- Nice paper. Definitely.



Rule 6: Don't give access to your code and data

You definitely cannot give others access to your raw data, especially if it doesn't exist! Fortunately, researchers have been avoiding sharing their data for decades:

- My data are not anonymized
- Data are available on my webpage
- You can contact me to (try to) get my data
- I cannot give it for free



Editorial of the New England Journal of Medicine (2016):

A second concern held by some is that a new class of research person will emerge – people who had nothing to do with the design and execution of the study but use another group’s data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited. There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as “research parasites.

–D.L. Longo and J.M. Drazen, *Data sharing*

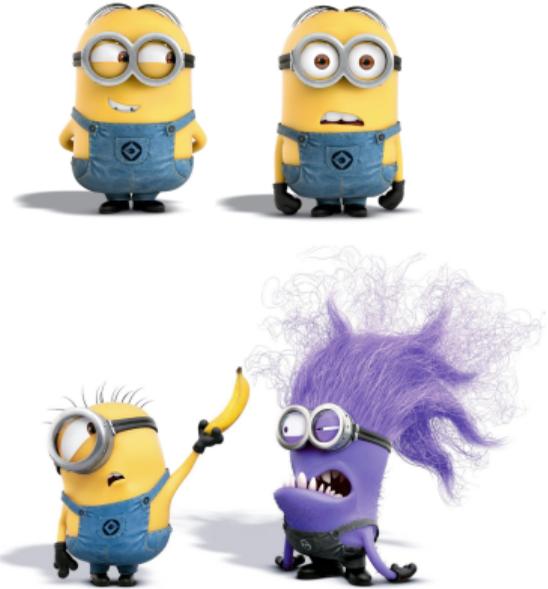
Anyway, there are **parasite awards** now.

Rule 7: Do not allow for replication outside your lab

It may surprise you, but some researchers may want to replicate your results using the methods explained in your article.

If people try to replicate your work and do not get the same results, you have a problem. They may insist on seeing your actual data and, if you refuse, you might be suspected of fraud or misconduct.

Thus, you can try to attach a “Do not replicate order” alongside your article.



Self-appointed data police are volunteering critiques of such personal ferocity and relentless frequency that they resemble a denial-of-service attack that crashes a website by sheer volume of traffic

...

*I have heard from graduate students opting out of academia, assistant professors afraid to come up for tenure, mid-career people wondering how to protect their labs, and senior faculty retiring early, all because of **methodological terrorism**.*

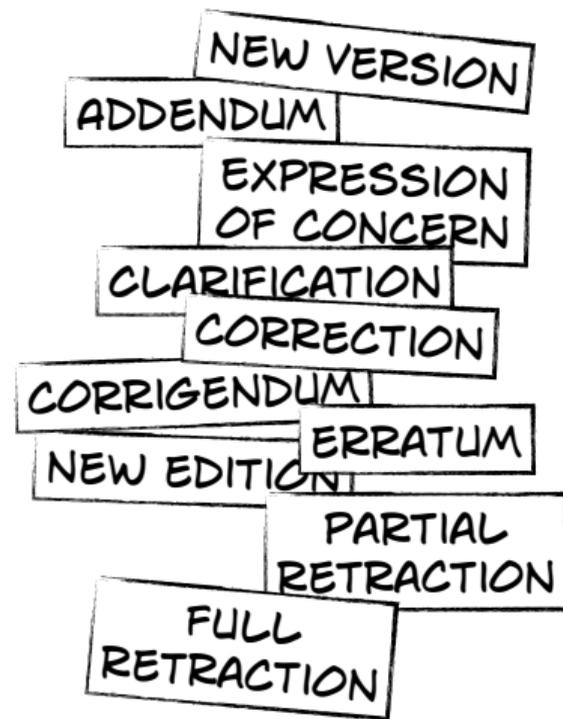
—Susan Finske, APS former president, 2016

Rule 8: Never, ever, retract your results

If you've been engaged in fraud, having your paper retracted is like an admission of guilt. It's something to be avoided.

Retraction could start with a simple comment to the editor who could decide to retract your paper without your consent.

Luckily for you, the whole retraction process can last for several years. Just the time necessary for everyone to forget the whole story.



Rule 9: Don't get caught. Deny if caught.

If you intend to persist in a rogue scientific career, you have to be aware that you're likely to get caught sooner or later. There is a set of simple rules to follow if you need to deny scientific misconduct.

I HAD NO
CHOICE

I'M A GIFT
AUTHORSHIP

I DIDN'T
KNOW I WAS
AUTHOR

I WAS NOT
AWARE

Ian Intern¹, John Bigname², John Dunno^{1,2}, Matt Supervisor^{1,*}



Rule 10: Be creative (for once)

Top ten retractions of 2017

- 107 papers from *Tumor Biology* (Elsevier) retracted (fake peer reviews)
- A paper about vaccine and autism was published twice and retracted twice
- Paper retracted (by authors) because the wrong transgenic mouse line was used.
- Jack Szostak (Nobel winner) retracted a paper from *Nature Chemistry* because his lab could not reproduce their own results.
- “Boom, Headshot!” paper retracted (fake data). One the co-authors stripped of her PhD.

The game is probably not worth the risks
(or is it ?)

Questions?