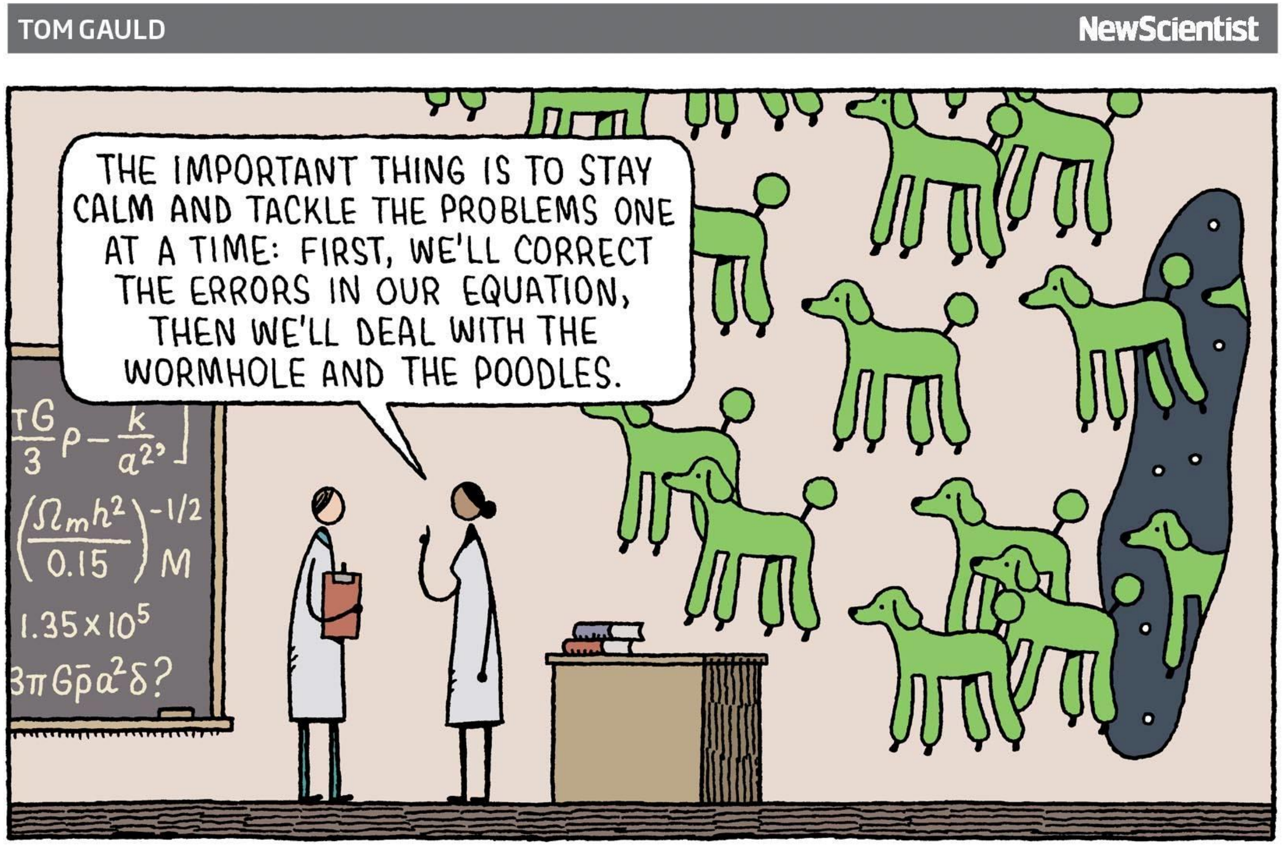
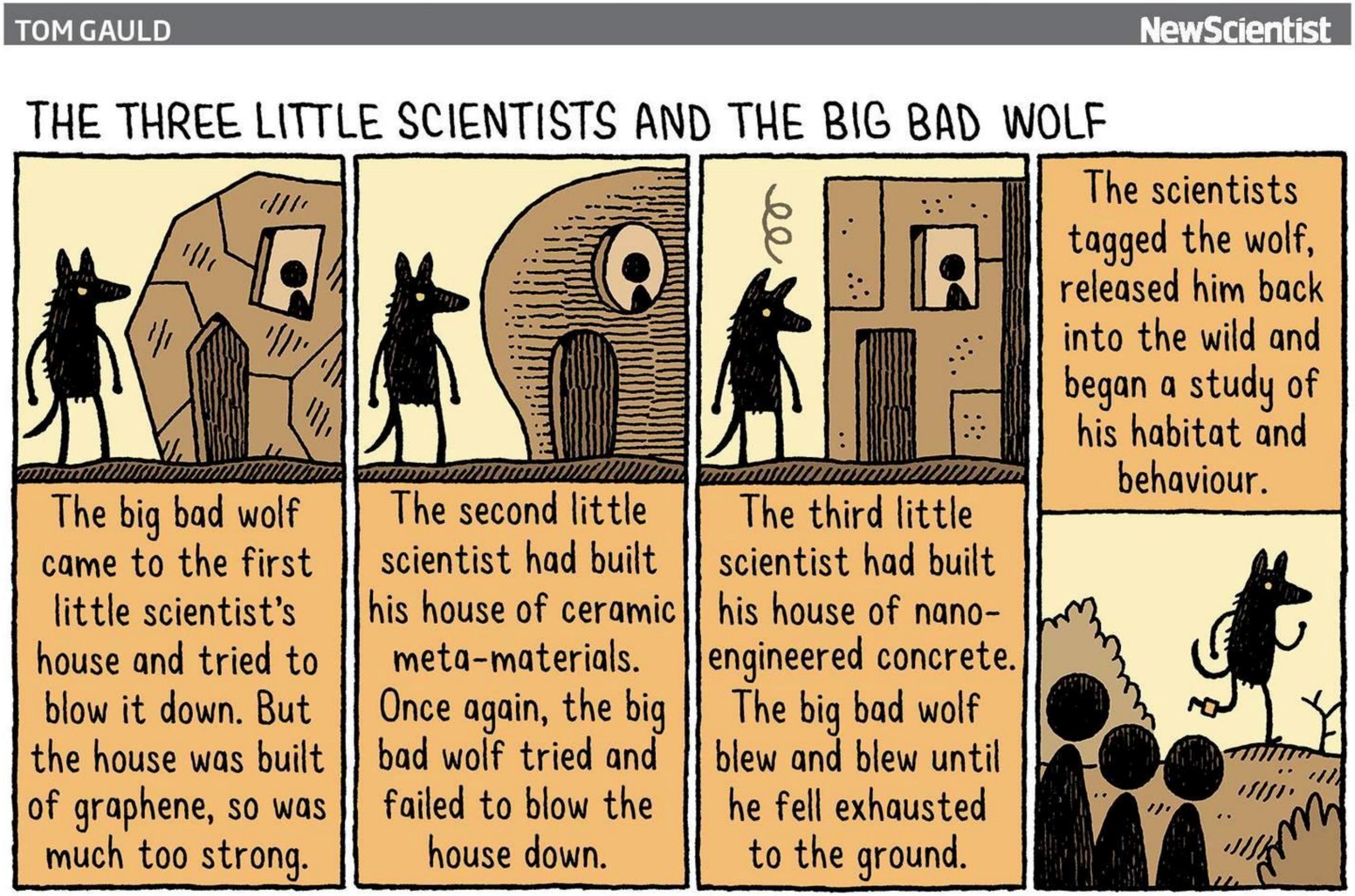
Reproducible Research

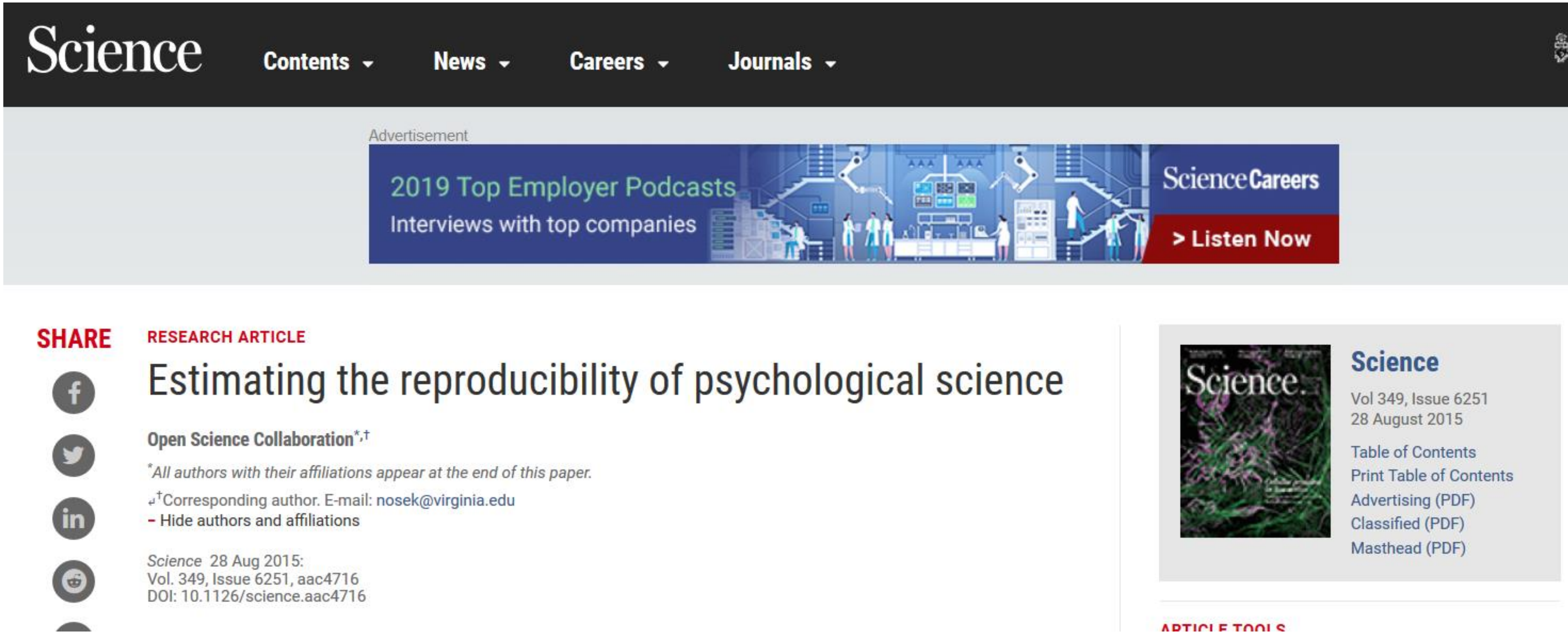

TOM GAULD
NewScientist
THE IMPORTANT THING IS TO STAY CALM AND TACKLE THE PROBLEMS ONE AT A TIME: FIRST, WE'LL CORRECT THE ERRORS IN OUR EQUATION, THEN WE'LL DEAL WITH THE WORMHOLE AND THE POODLES.
$\frac{\pi G}{3}\rho - \frac{k}{a^2}$]
$\left(\frac{\Omega_m h^2}{0.15}\right)^{-1/2} M$
1.35×10^5
$3\pi G\bar{\rho}a^2\delta$?

SML310: Research Projects in Data Science, Fall 2019

Michael Guerzhoy

(Unrelatedly)


TOM GAULD
NewScientist
THE THREE LITTLE SCIENTISTS AND THE BIG BAD WOLF
The big bad wolf came to the first little scientist's house and tried to blow it down. But the house was built of graphene, so was much too strong.
The second little scientist had built his house of ceramic meta-materials. Once again, the big bad wolf tried and failed to blow the house down.
The third little scientist had built his house of nano-engineered concrete. The big bad wolf blew and blew until he fell exhausted to the ground.
The scientists tagged the wolf, released him back into the wild and began a study of his habitat and behaviour.



Science
Contents
News
Careers
Journals

Advertisement
2019 Top Employer Podcasts
Interviews with top companies
Science Careers
> Listen Now

SHARE

RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Open Science Collaboration[*,†]

[*] All authors with their affiliations appear at the end of this paper.

[†] Corresponding author. E-mail: nosek@virginia.edu

- Hide authors and affiliations

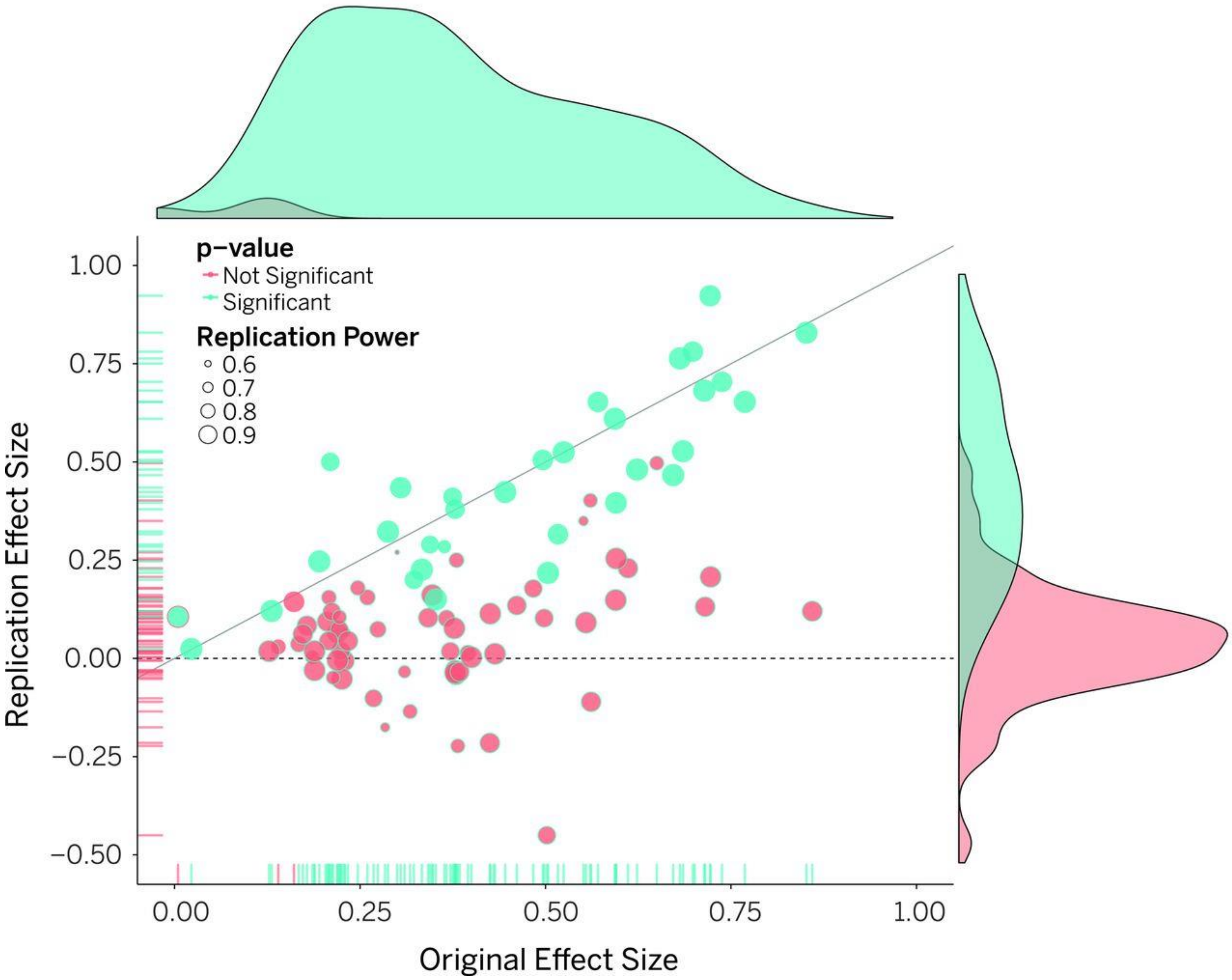
Science 28 Aug 2015:
Vol. 349, Issue 6251, aac4716
DOI: 10.1126/science.aac4716

Science
Vol 349, Issue 6251
28 August 2015

Table of Contents
Print Table of Contents
Advertising (PDF)
Classified (PDF)
Masthead (PDF)
ARTICLE TOOLS

Open Science Collaboration study

- Replications of 100 experimental and correlational studies
- “The mean effect size (r) of the replication effects (M_r = 0.197, SD = 0.257) was half the magnitude of the mean effect size of the original effects (M_r = 0.403, SD = 0.188)”
- “97% of original studies had significant results ($P < .05$). 36% percent of replications had significant results”
- “47% of original effect sizes were in the 95% confidence interval of the replication effect size”
- “39% of effects were subjectively rated to have replicated the original result”
- “if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects”


p−value
Not Significant
Significant
Replication Power
0.6
0.7
0.8
0.9
Replication Effect Size
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
0.00
0.25
0.50
0.75
1.00
Original Effect Size

Distribution of p-values

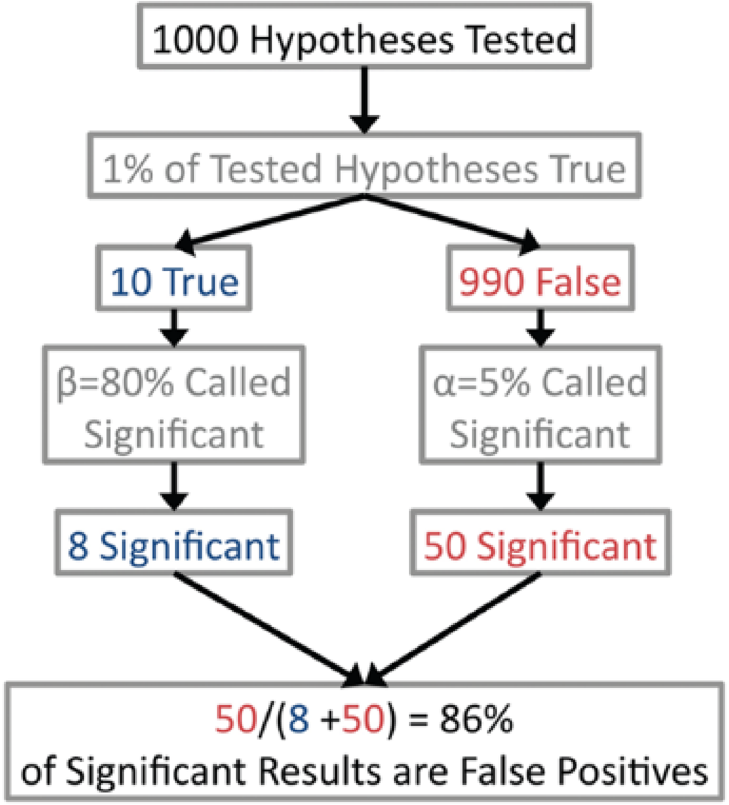
- P-value: the probability of observing an observation as extreme or more extreme than what was actually observed, assuming the null-hypothesis is true
- If the Null-Hypothesis is true, what is the probability that the p-value is smaller than 0.3?

Distribution of p-values

- P-value: the probability of observing an observation as extreme or more extreme than what was actually observed, assuming the null-hypothesis is true
- If the Null-Hypothesis is true, what is he probability that the p-value is smaller than 0.3?
  - 30% of the time, will get a value such that we'd get an even more extreme value 30% or less.

Distribution of p-value

- Some null-hypotheses are true, and some are false


1000 Hypotheses Tested
1% of Tested Hypotheses True
10 True
990 False
β=80% Called Significant
α=5% Called Significant
8 Significant
50 Significant
50/(8 +50) = 86%
of Significant Results are False Positives

If most of the hypotheses tested are actually false, most of the positive findings are false positives

What percentage of hypotheses being test is false?

- Suppose the fraction of false null hypotheses is π_0
- Consider the distribution of the p-values for “positive” results ($p < 0.05$)
- The probability of seeing a particular p-value can be modelled as

$$f(p|a,b,\pi_0) = \pi_0 Unif(0, 0.05) + (1-\pi) tBeta(a, b; 0.05)$$

- $tBeta(a, b; 0.05)$ is the Beta distribution, truncated at 0.05
- Can now infer π_o using a version of Maximum Likelihood

Intuition

- What kind of distribution of p-values would we expect if the null-hypothesis is true?
- What kind of distribution of p-values would we expect for the entire field, with some true hypotheses and some false hypotheses?

So what’s the false discovery rate?

- Original estimate: 15%
  - The model does *not* include bad faith or fraud
- Authors then acknowledged it may be higher