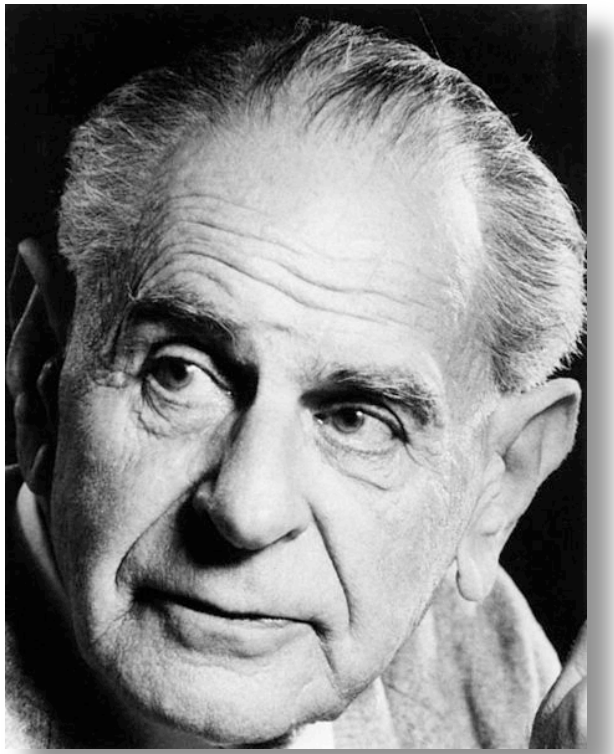




Scientific Reasoning

Sam Schwarzkopf

www.neuroneurotic.net

The “Psi” hypothesis

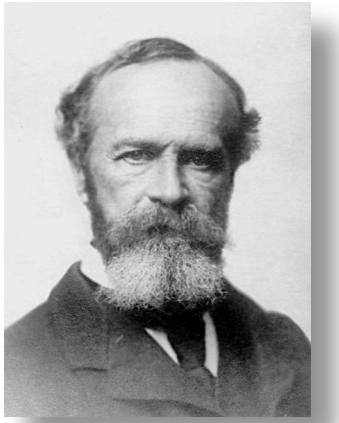
‘…anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms.’

Bem (2011). J Pers Soc Psychol

Includes purported phenomena like precognition, telepathy, clairvoyance, and psychokinesis.

Frequently linked with time-symmetry in physics and quantum mechanics.

The “Psi” hypothesis



In some form or other this has been part of psychology since the early days.

Was subject to very serious study, including projects by the Central Intelligence Agency.

Several notable scientists today are Psi proponents.

Parsimony (“Occam’s Razor”)

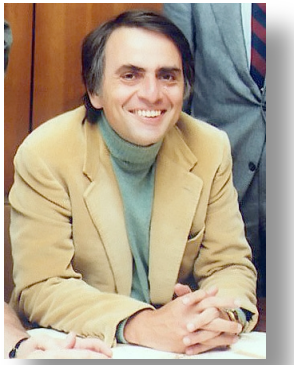
The explanation that requires the *least assumptions* is *probably correct*.

Parsimony ("Occam's Razor")

The explanation that requires the *least assumptions* is *probably correct*.

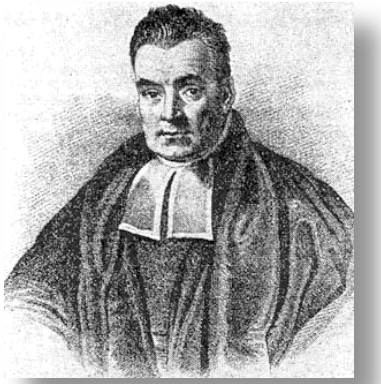
Extraordinary claims require extraordinary evidence

More alternative explanations may suggest a *more extraordinary* claim, while more *specific results* suggest *weaker evidence*.



Does the result meet our predictions?

How close is our observed *effect size* to what we can *plausibly expect* from this hypothesis?



William of Ockham picture by Moscarlop (https://en.wikipedia.org/wiki/File:William_of_Ockham.png)

Scientists are not immune

Implausible ideas and logical fallacies are wide-spread. They may underlie conspiracy theories.

These are cognitive biases. And as human beings scientists are not immune to them.

It is very easy to develop a blind-spot where your favourite hypothesis is concerned…

“Feeling the future…”

Controversial study in a psychology journal

Bem (2011). J Pers Soc Psychol

Presented 9 experiments that test the hypothesis that humans can predict or sense the future.

All experiments are largely classical psychology experiments in which sequence has been reversed.

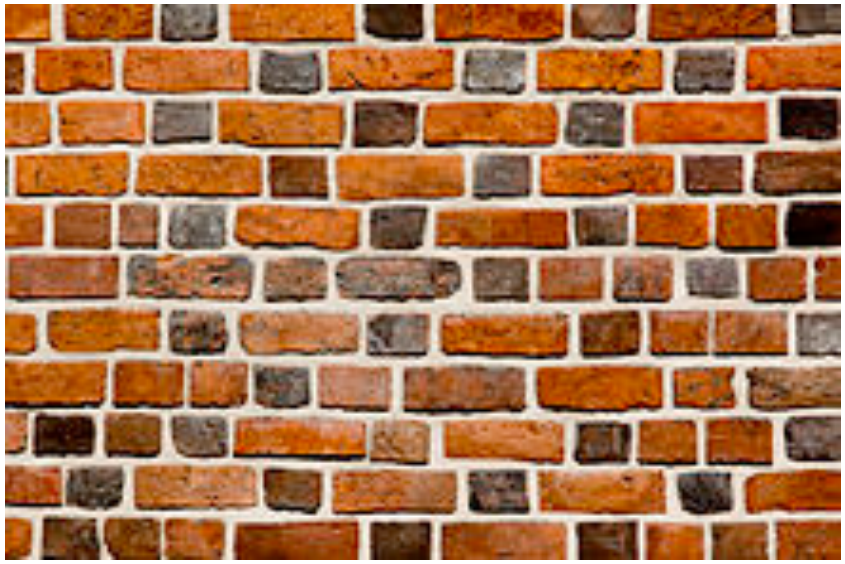
One (but by far not the only!) impetus for the current crisis in psychology research.

Bem’s experiment 1

100 participants (50 female, 50 male) each performed 36 trials of the following task…

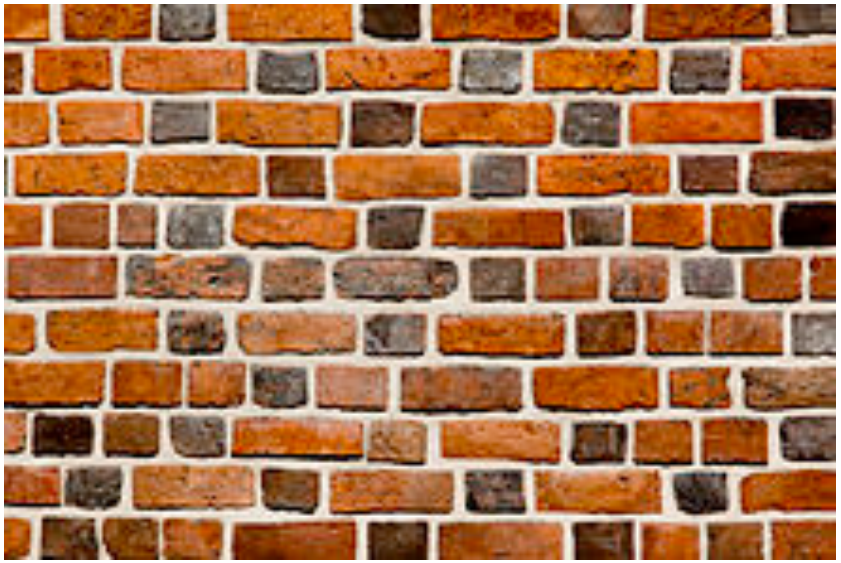
“Click on the curtain you feel has a picture behind it…”

But in truth the computer only decides where the picture is *after* the participant’s response.

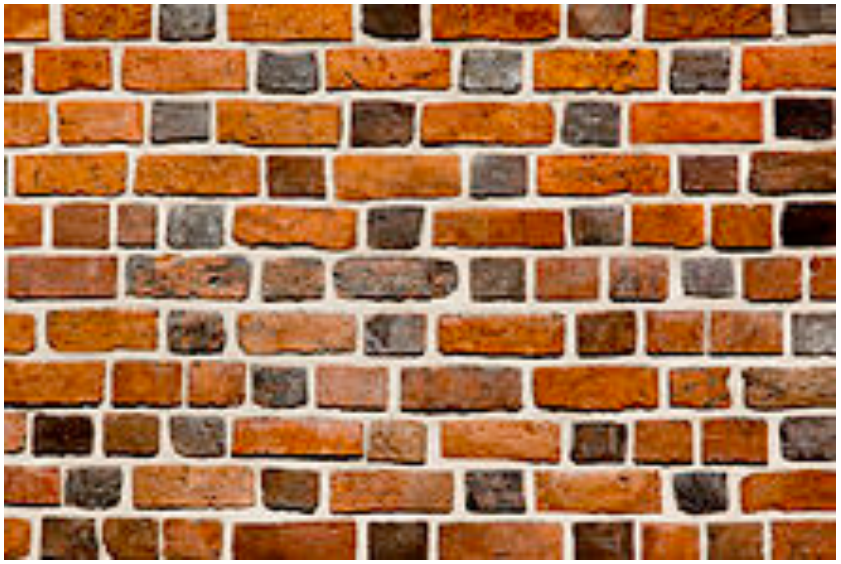


12 trials contained “positive” (erotic) images

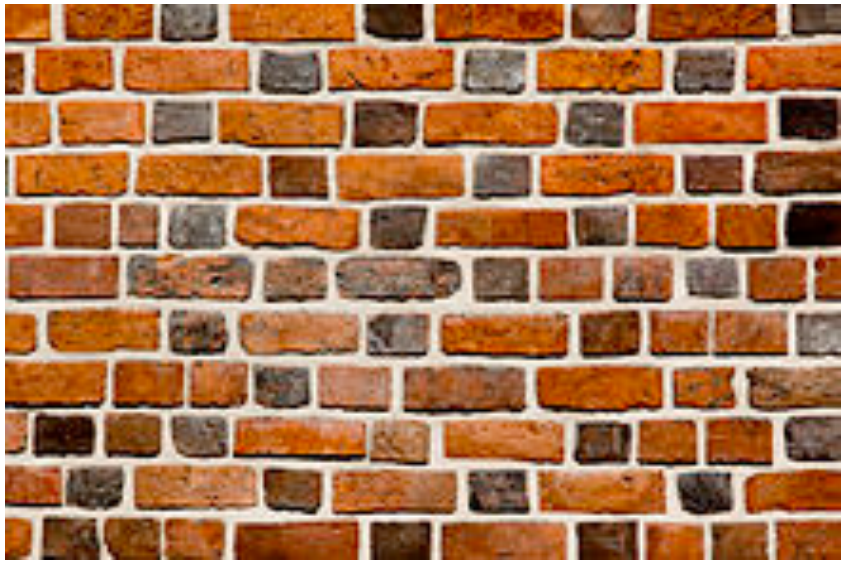
(with humans not turtles…)



12 trials contained negative (horrifying, scary) images



12 trials contained neutral images



The final 60 participants instead performed 18 trials with erotic pictures and 18 with non-erotic positive

Bem's experiment 1: Results

On average participants were correct on 53.1% of erotic trials, significantly above chance:

t(99) = 2.51, p = 0.014

They only appeared to be guessing on the non-erotic trials. Accuracy was 49.8%, which was not significantly different from chance:

t(99) = 0.15, p = 0.56

Is this sufficient evidence that precognition exists?

How likely is it that precognition exists?

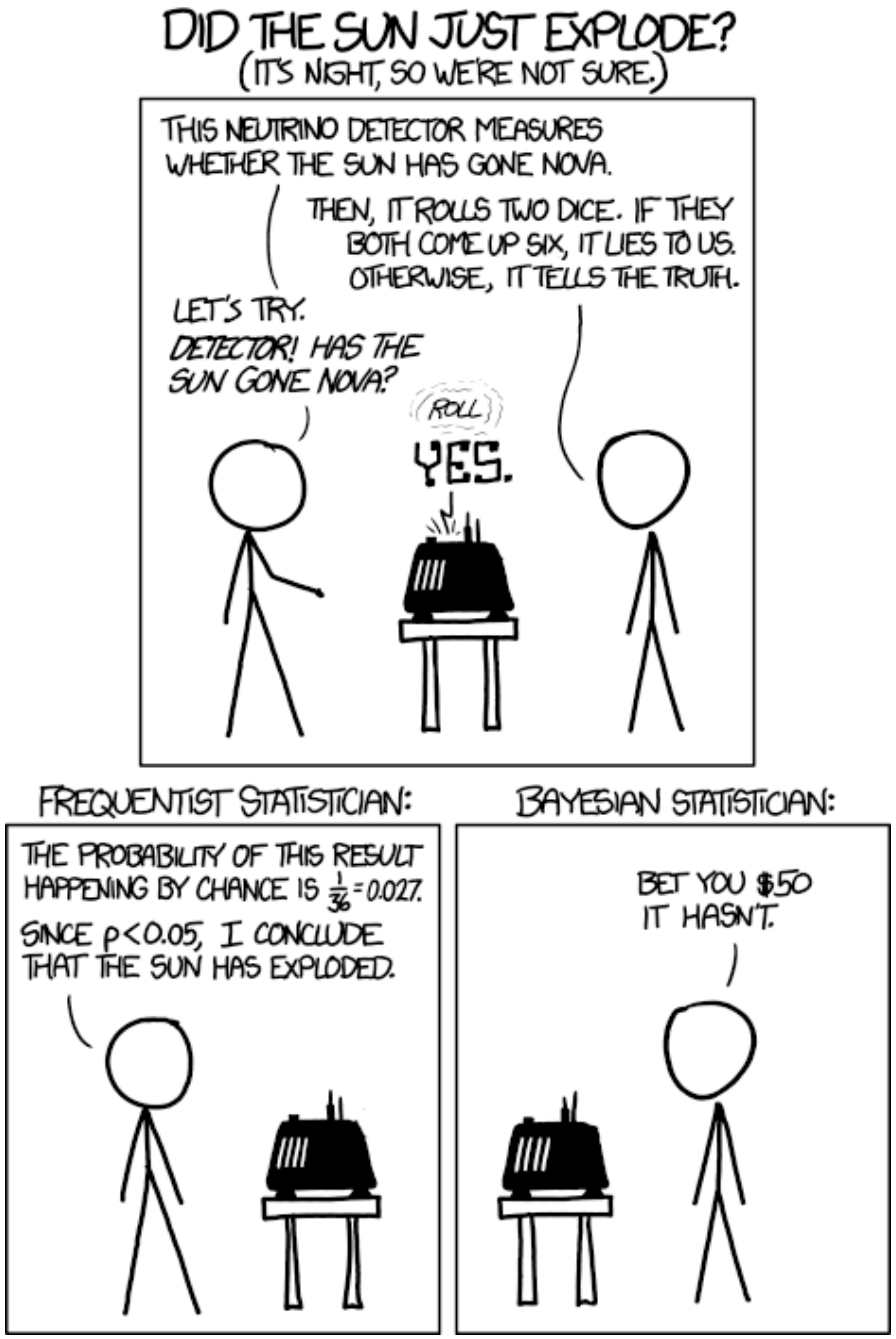
The base-rate fallacy

"Statistically significant" means a result as or more extreme as our observation is unlikely *under the assumption that there is no true effect*.

This does not account for the prior probability that the hypothesised effect is true.

Thus it also does not account for the likelihood of observing the result assuming the hypothesis is true.

$$P(H_1|D) \neq P(D|H_1)$$



https://xkcd.com/1132/

Priming behaviour

Influential theory that posits simple stimuli – even subliminal ones – can affect complex behaviour.

People primed with “rudeness” become more impolite (i.e. they interrupt more quickly)
Bargh et al. (1996) J Pers Soc Psychol

People primed with “professor” become smarter than those primed with “football hooligan”
Dijksterhuis & van Knippenberg (1998) J Pers Soc Psychol

People “primed with money … play alone, work alone” and keep “more physical distance” Vohs et al. (2009) Science

More examples of priming…

Holding hot or cold beverage alters peoples' "social proximity" and "relational focus"
IJzerman & Semin (2009) Psych Sci

Cleanliness or washing your hands reduces the severity of moral judgements Schnall et al. (2008) Psych Sci

Turning a crank clockwise makes you prefer new things
Topolinski & Sparenberg (2012) Soc Psych Pers Sci

Being outside a box makes you more creative
Leung et al. (2012) Psych Sci

A tiny US flag in corner of the screen makes you vote Republican – 8 months in the future! Carter et al. (2011) Psych Sci

How can we know that a scientific finding is “true”?

Any single result is never conclusive evidence…

…regardless of p-value, statistical power, or impact factor

Science is the search for *regularities in nature*

Only replication can reveal regularities

Direct replication shows reliability

Conceptual replication shows generalisability

Independent replication more solid evidence than internal replication

But what does a failure to replicate mean?

What can a failed replication tell us?

1. The original finding could simply be false.

2. Something may be different in the replication.

What can a failed replication tell us?

“Angry birds … nothing in their heads...”

“…hand-wringing ... largely pointless...”

“…shameless little bullies...”

“…replication police...”

“…methodological terrorism...”

What can a failed replication tell us?

“I'm up for tenure next year; I can't be on a paper that argues against my effect.”

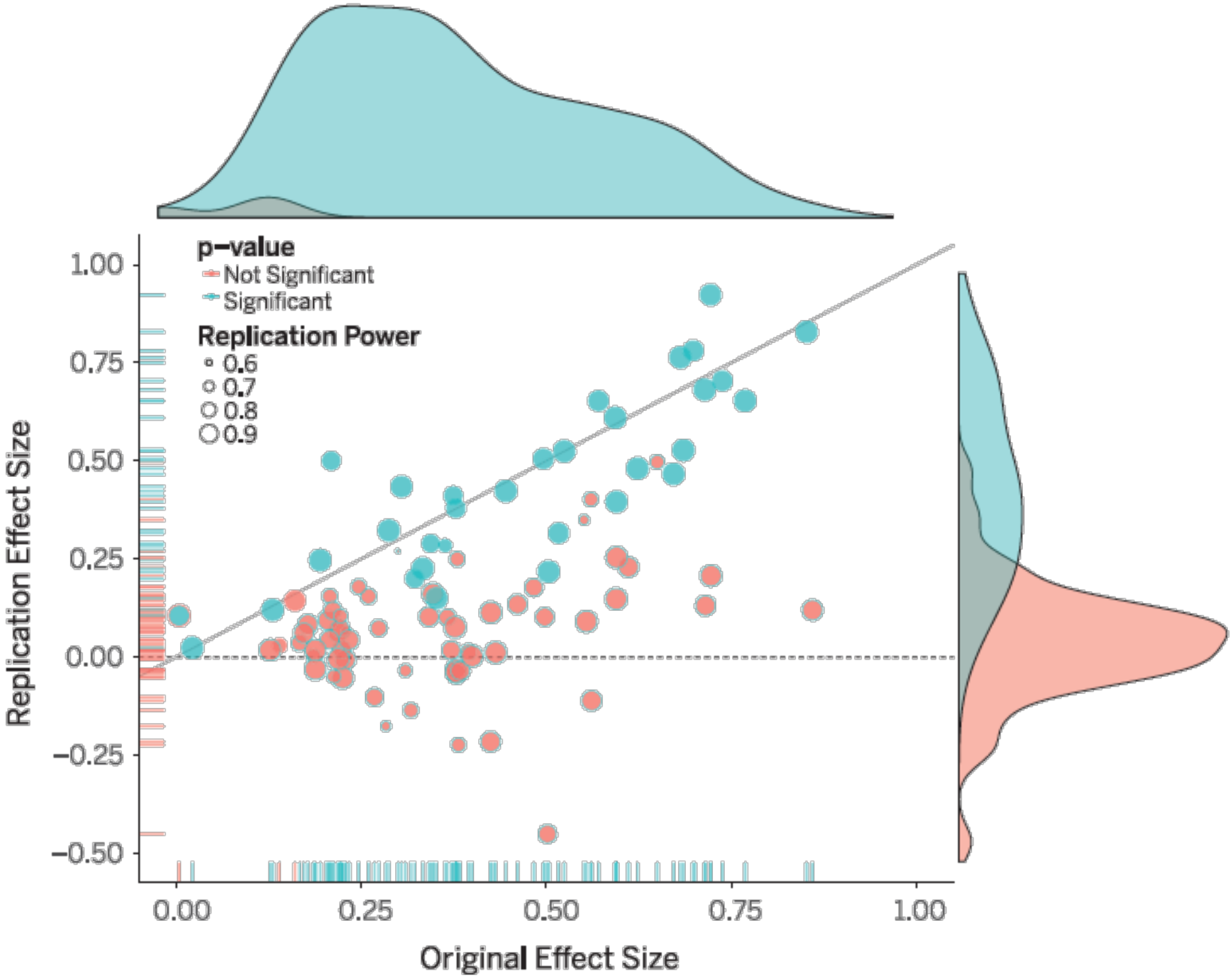
Estimating the replicability* of psychology research

Large-scale replication attempt by the Open Science Collaboration (2015).

100 findings from psychology literature. Each was attempted to be replicated by one lab.

The outcome isn't exactly encouraging…

**They called it "reproducibility" but they mean "replicability": redoing the same experiment with same methods to collect new data.*


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

Straight-up replications don't show us all that much

Large number of replication failures is an *indictment of robustness* of science. However:

Failure to replicate could be due to differences between different studies…

A successful replication could just replicate the same mistakes as the original…

A good replication attempt aims to discover something new

Seek to replicate the original effect but at the same time test some new factor.

You may still fail to replicate the original effect – but at least your intentions were clear then.

If you don't test a new factor, *at the very least try to improve* upon original methods in some way.
(In fact, one could call this testing a new factor)

Bargh et al (1996): “Elderly priming”

Participants were given cover story that the study was about language processing…

Participants had to produce a 4-word sentence from sets of 5 words (30 trials).

For one group of participants the stimuli contained words related to old age.

Bargh et al (1996): “Elderly priming”

grey

bitter

rigid

lonely

selfish

helpless

Florida

withdraw

forgetful

gullible

cautious

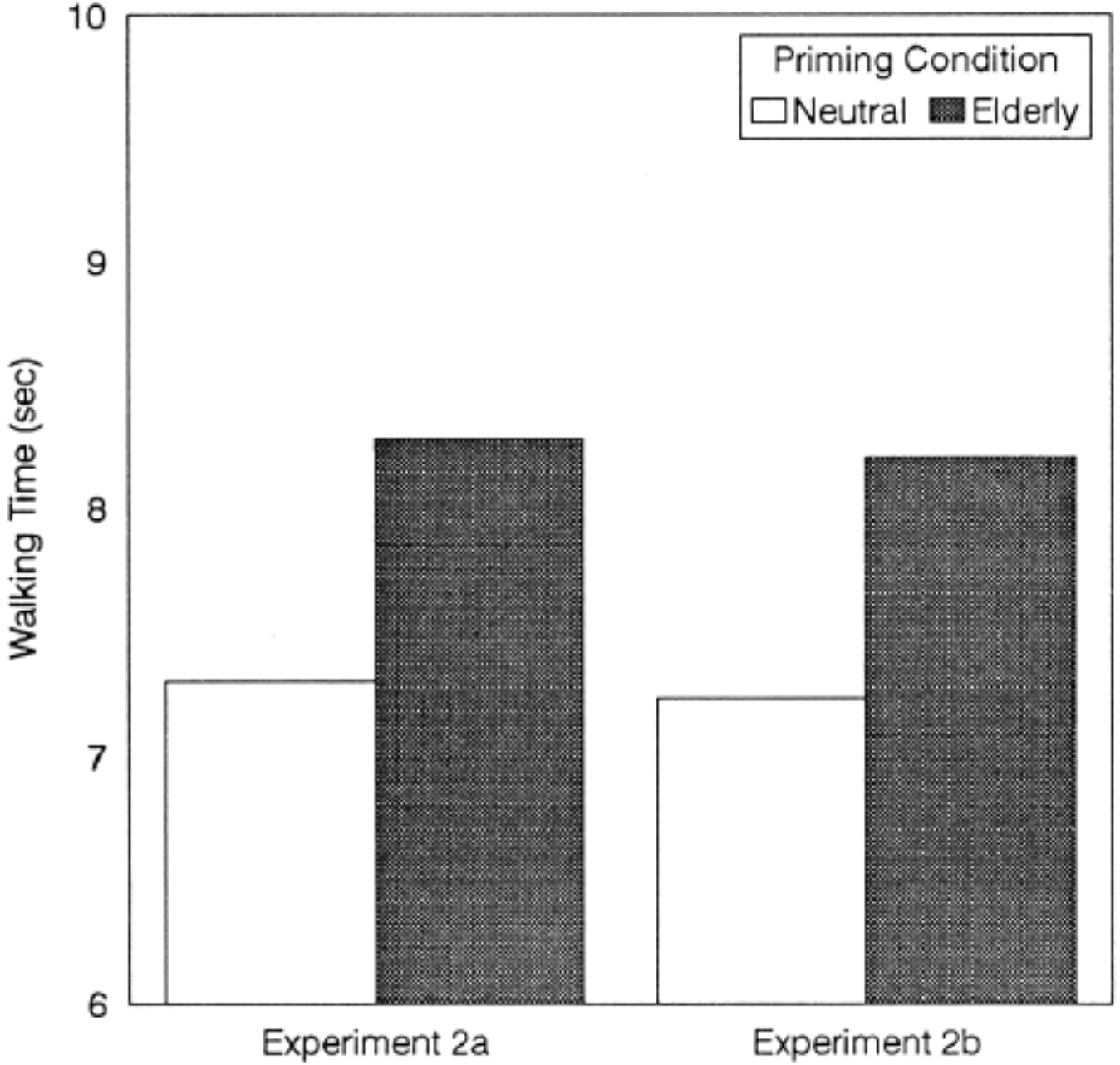
Bargh et al (1996): “Elderly priming”

Participants were given cover story that the study was about language processing…

Participants had to produce a 4-word sentence from sets of 5 words (30 trials).

For one group of participants the stimuli contained words related to old age.

People primed with “old age” walked more slowly as they left the lab.


Priming Condition
Neutral
Elderly
Walking Time (sec)
10
9
8
7
6
Experiment 2a
Experiment 2b

Bargh et al (1996): “Elderly priming”

grey

bitter

rigid

lonely

selfish

helpless

Florida

withdraw

forgetful

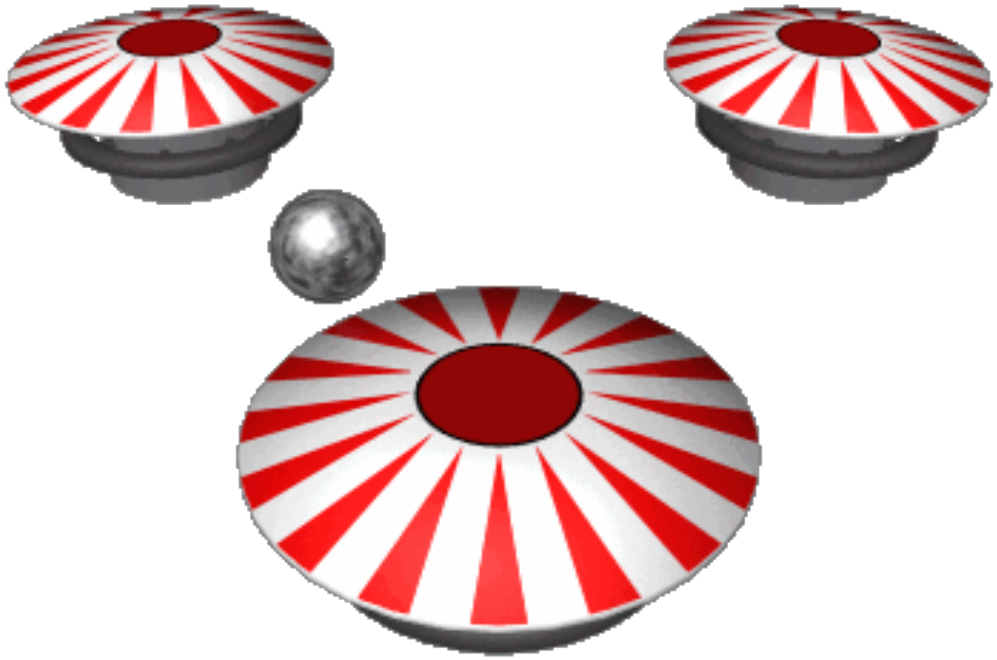
gullible

cautious


STOP

Humans are dirty test tubes

How plausible is social priming?



Effect sizes are inconsistent with the multitude of factors that affect each participant differently…

Problems with “elderly priming”

Walking speed was measured by confederate with a stopwatch.

Unclear if participants were really unaware of the manipulation.

Also unclear whether experimenter was unaware of the experimental condition.

Replication by Doyen et al. (2012)

Walking speed was measured more accurately by infrared sensors.

More thorough debriefing of awareness.

Basic replication failed to show any differences in walking speed.

Second replication suggested that experimenters' expectation influences participants' walking speed.

A good, *successful* replication can *falsify* a hypothesis

A successful replication that shows the effect occurred for another reason than originally hypothesised.

In contrast, a failed replication certainly argues against the robustness of an effect but cannot *falsify* it.

But every replication result is in itself a new hypothesis to be falsified/disproven…

Problems with Doyen replication

Original study was in 1996, the replication in 2012.

Original study conducted in US, replication in Belgium.

Original stimuli were in English, replication stimuli were in (Belgian) French.

Experimenter influence on participant is a vague hypothesis. How does that happen?

Defences by priming researchers

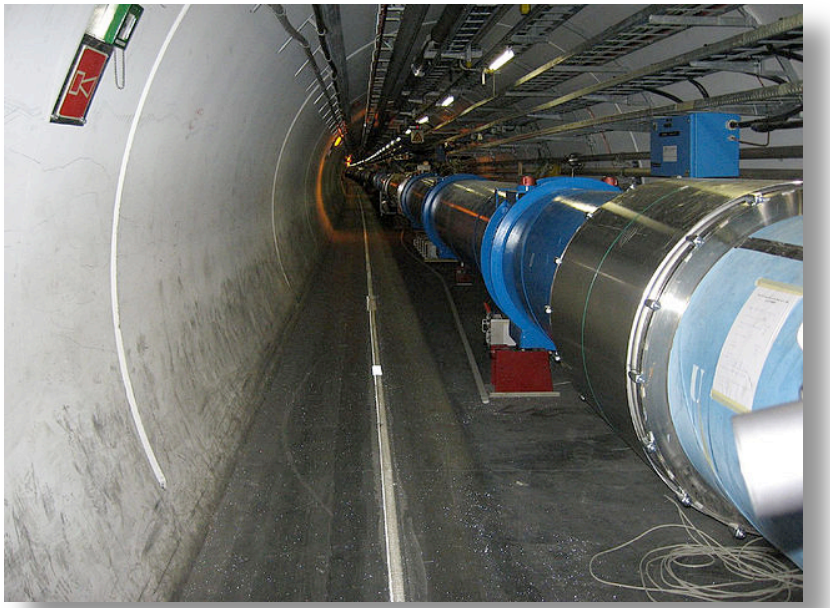
There are “hidden moderators,” e.g. priming with “professor” makes you smarter but priming with “Einstein” makes you less intelligent…

-> *Complex & specific effect = Weaker evidence*

Experienced priming researchers just “know” better how to produce these effects than “replicators”

Problems with expertise defence

Skill and experience are doubtless important –
but administering a questionnaire or psychology task also is not really the same as running the LHC



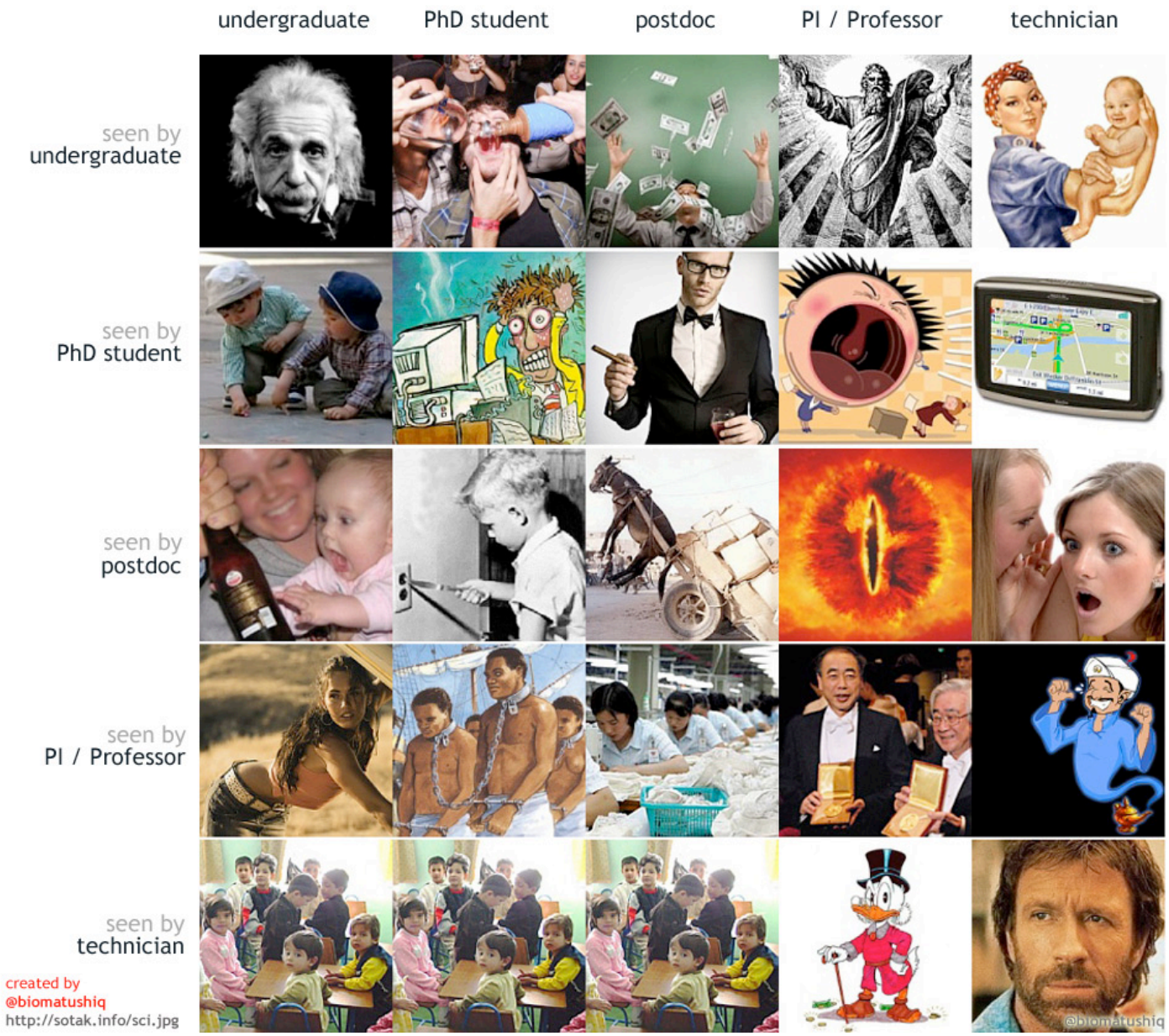
Problems with expertise defence

Skill and experience are important –
but administering a questionnaire or psychology task also is not really the same as running the LHC

Post-hoc hypotheses are fine –
but you must test them experimentally

In most cases the “experienced researchers” are not even the ones running the experiments!

How people in science see each other



Strong inference

Don’t just test one hypothesis (”my effect”) but contrast two or more.

Ideally, positive evidence for one hypothesis refutes the other one.

Platt (1964) *Science*

Summary

Science doesn't prove hypotheses but compares the strength of evidence for different hypotheses.

Always ask how likely it is that a given hypothesis is true.

Experiments should test how close the results are to the predictions made by different hypotheses.

When faced with multiple alternative explanations, keep an eye on the least complex one.

"Give the null hypothesis a chance" Alcock (2003). J Consc Stud

Think about what could convince you that your favourite hypothesis is untrue.


Welcome to the
Sea of Doubt!