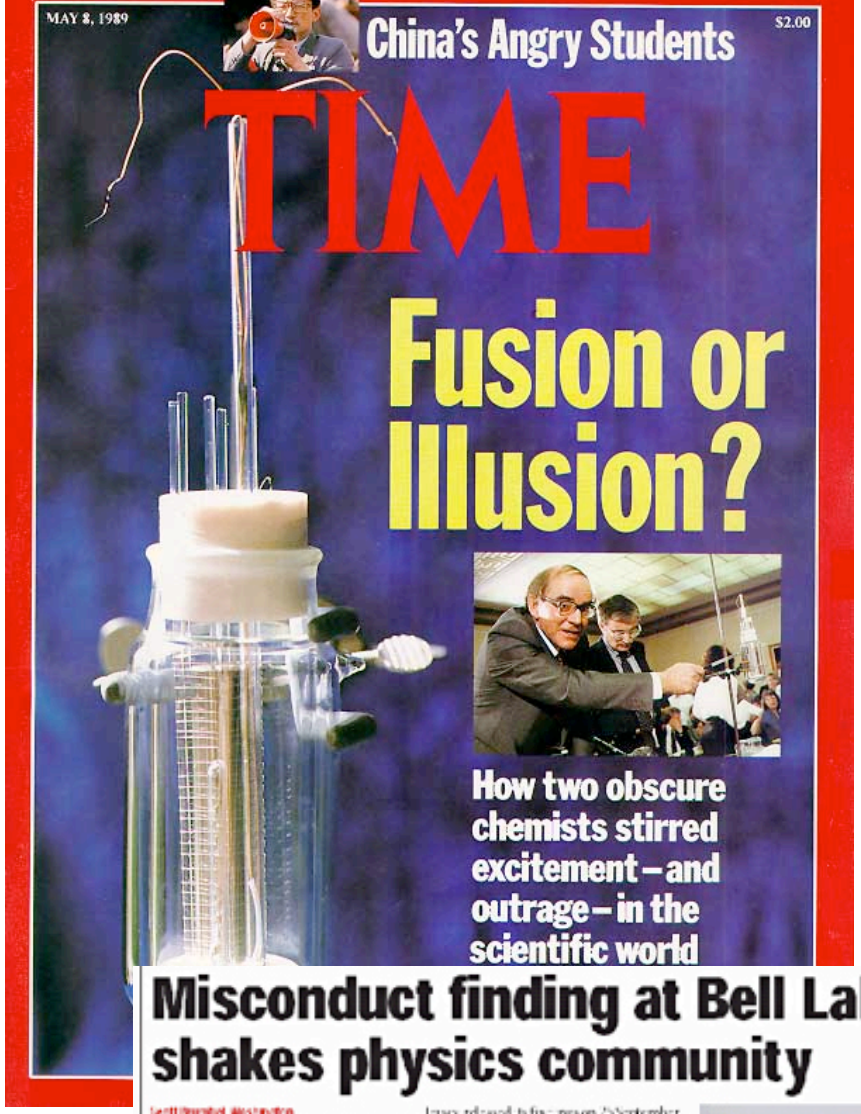
Journée d'étude du Cogmaster
Good (and less good) lab practices

Saturday 25 octobre 2014

- 9h30-10h30: overview of lab practices
- 11h30-15h30 Lab Notebook
  - 11h35-11h55: summarizing/clarifying the notebook
  - 11h55-12h20: questions/discussion
  - LUNCH BREAK
  - 13h30-14h30: error finding
  - 14h30-15h30: questions/discussion
- 15h30-16h30 Wrapping up (& cleaning up)



Published online 21 October 2009 | *Nature* **461**, 1035 (2009) | doi:10.1038/4611035a

News

Hwang verdict imminent

Disgraced stem-cell researcher may face time in jail.

David Cyranoski

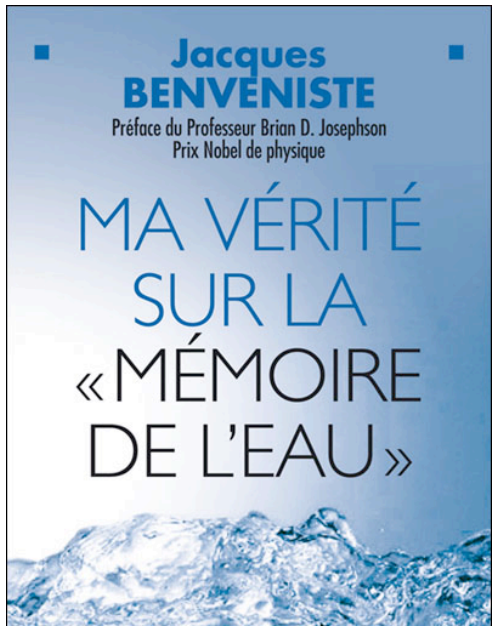
Despite his research being exposed as fraudulent and unethical almost four years ago, the career of South Korean cloner Woo Suk Hwang has thrived. He has established a research institute, laid claim to a set of human-cloning patents, received a scientific excellence award, published a handful of papers and entered into a collaboration with a powerful provincial government.

Hwang is charged with fraud.

CHUNG SUNG-JUN/GETTY

Misconduct finding at Bell Labs shakes physics community

Is a bell tolling for B



- Cognition hall of f/sh/ame

Mark Hauser
(cognitive pschology, Harvard University)
Suspended, 2011

Diederik Stapel
(social pschology, Tilburg University)
Convicted of Fraud, 2012
fraud in 55 papers

Google Scholar
nb of publications: 142
nb of citations: 4312
h index: 38

Questions: how did they end up doing this? (mechanisms)
- misbehavior from the start vs gradual change?
- conscious vs unconscious (cognitive dissonance reduction)?

what were the motivations?
- external forces (publication pressures) vs internal motivations ?

historical precedents

- *One can applaud the lucky gambler; but when he is lucky again tomorrow, and the next day, and the following day, one is entitled to become a little suspicious*" [5].
- *results as close as or closer to expectations as the ones reported by Mendel would occur in only 1 out of 33,000 replications* [6, p. 921]

Gregor Johann Mendel

The data seemed to be a little bit too ‘clean’. but was is deliberate ‘cheating’?

credibility of scientific research

The gray area

- data collection: experimenter biases
- data analysis: data massaging (today's focus)
- write up: biased rendering of the literature, misreporting of data, unconscious plagiarism, authoring issues
- reviewing & publication biases: null results/non replication avoidance, forced citations

The gray area

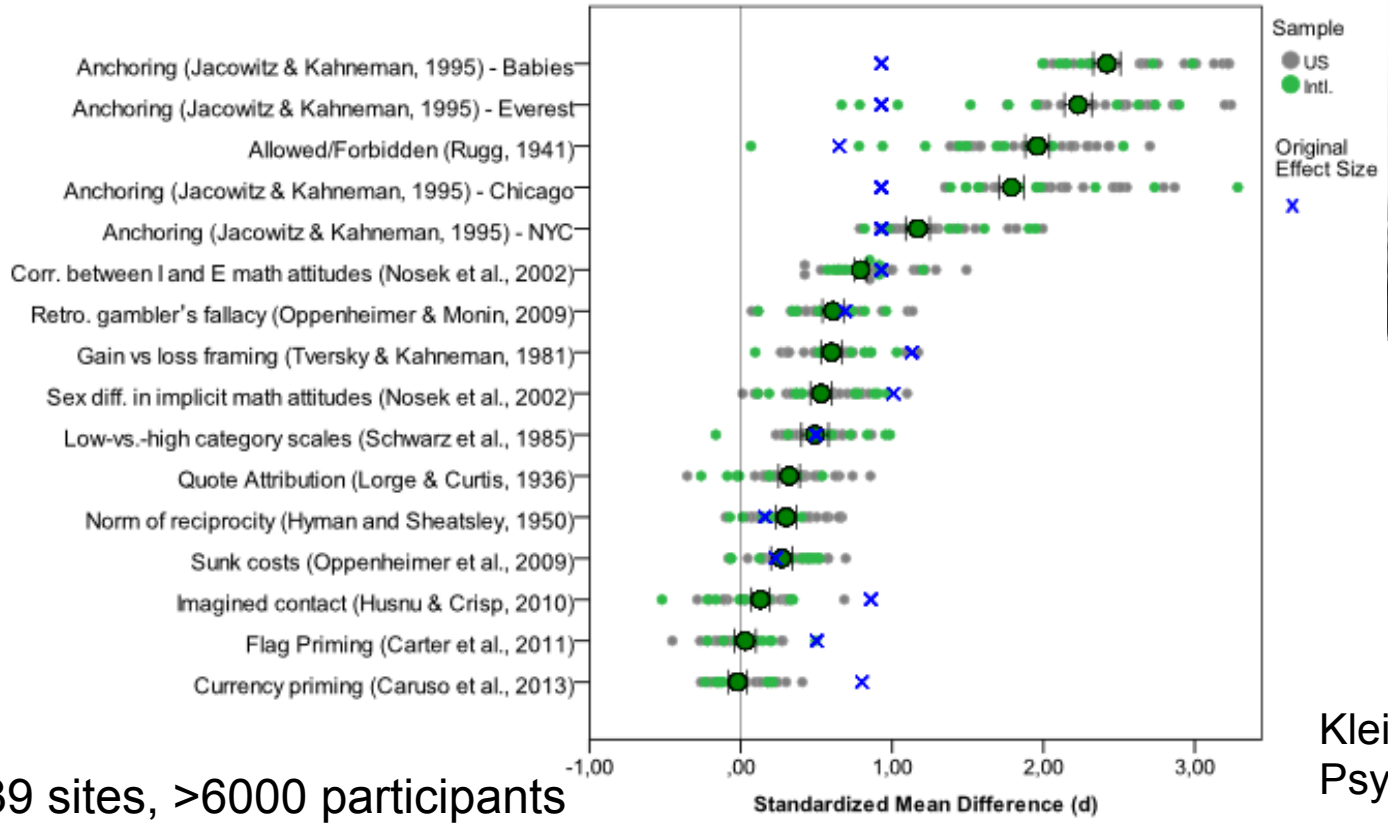
- data collection: experimenter biases
- data analysis: data massaging (this morning)
- write up: biased rendering of the literature, misreporting of data, unconscious plagiarism, authoring issues
- reviewing & publication biases: null results/non replication avoidance, forced citations

→ *Scientific bubbles, the decline effect, zombie beliefs*

the decline effect

- Lehrer (2010): effect sizes decline with time
  - Q1: explain the decline effect in terms of cognitive or other biases
  - Q2: what does it say about science

the replicability crisis



Schnall et al 2008

Klein et al. (2014). Social Psychology.

39 sites, >6000 participants

- 3/16 do not replicate (18%)
- most original studies outside of 99% confidence interval
- no effect of on-line vs lab experiments

see osf.io

ongoing: many labs 2

Possible “bubble” mechanisms

- data collection: experimenter biases
- data analysis: data massaging (this morning)
- write up: biased rendering of the literature, misreporting of data, unconscious plagiarism, authoring issues
- reviewing & publication biases: null results/non replication avoidance, forced citations

experimenter biases

- Rosenthal & Lawson (1964)
  - bright vs dull rats

Experiment	(N)	Bright	Dull	(N)	Mann-Whitney U One-tailed p	Rank Correlation with preceding experiment
I. Magazine Training	(7)	4·4	5·8	(5)	0·13	
II. Operant Acquisition	(8)	4·3	6·2	(5)	0·09	0.25
III. A. Extinction	(6)	4·2	5·8	(5)	0·12	0·08
B. Spontaneous Recovery	(8)	4·6	5·0	(6)	0·48	0·25
IV. Secondary Reinforcement	(6)	4·7	5·5	(4)	0·17	0·37
V. Stimulus Discrimination	(8)	4·0	6·3	(6)	0·008	0·38
VI. Stimulus Generalization	(7)	4·3	5·8	(4)	0·02	0·59 ($p < 0·05$, one-tailed)
VII. Response Chaining	(8)	5·8	3·8	(5)	0·17 (two-tailed)	0·45
Means	(7·3)	4·5	5·5	(5·0)	0·015	0·35 ($p < 0·005$, one-tailed)

possible mechanism?

	Bright	Dull	t	$p < 0{\cdot}10$ (two-tailed)
I. Satisfaction with Experiment	9·1	6·6	4·40	0·001
F. Much Handling of *S*s				
1. Before each experiment	−1·2	−2·3	<1	
2. After each experiment	−0·8	−3·3	1·17	
[3. Total Handling	−1·0	−2·8	3·34	0·09]
G. Much Watching of *S*s (new scale)	9·3	8·3	2·16	0·06
H. Much Talking to *S*s (new scale)	−3·1	0·7	2·41	0·04

Examples in humans

- influence of experimenter on subject's reaction times, verbal memory, psychophysical judgments, IQ scores, Rorschach responses, athletic performance, … (Rosenthal, R. & Rubin, D.B. (1978). Behavioral and Brain Sciences, 3, 377-386)
- secondary effects: influence of experimenter on implicit influences on subjects

Consequences outside the lab

- favoritism in classrooms (Rosenthal, R. and Jacobson, L. (1968). Pygmalion in the Classroom (Holt, Rinehart and Winston, New York)
- gender biases (maths vs art)

- punchline
  - experimenter biases are subtle, unconscious, but can quite systematically influence experimental results
  - it is important to acknowledge that they exist in order to correct for them

does it affect my research?

- do I compare several groups of subjects? (eg patients versus controls, experimental group vs control group, before training vs after training)
- how to deal with it?

does it affect my research?

- do I compare several groups of subjects? (eg patients versus controls, experimental group vs control group, before training vs after training)
- how to deal with it?
  - double blind protocols
    - (ie, the participant is not aware of the research question, the experimenter does not know which condition the subject is in)
  - fully computerized protocols (the computer assigns the participant to a group)
  - internal controls (the groups of subjects are also tested on a neutral condition which should be the same in both groups)
- do I compare conditions within subjects?
  - fully computerized protocols (the computer decides which condition is presented when, and the experimenter is either blind or not present)

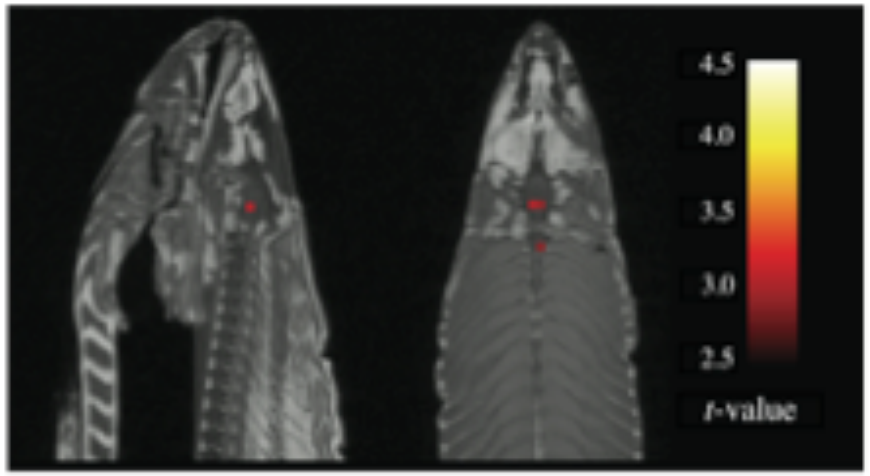
Possible “bubble” mechanisms

- data collection: experimenter biases
- data analysis: data massaging (this morning)
- write up: biased rendering of the literature, misreporting of data, unconscious plagiarism, authoring issues
- reviewing & publication biases: null results/non replication avoidance, forced citations

the dead salmon effect

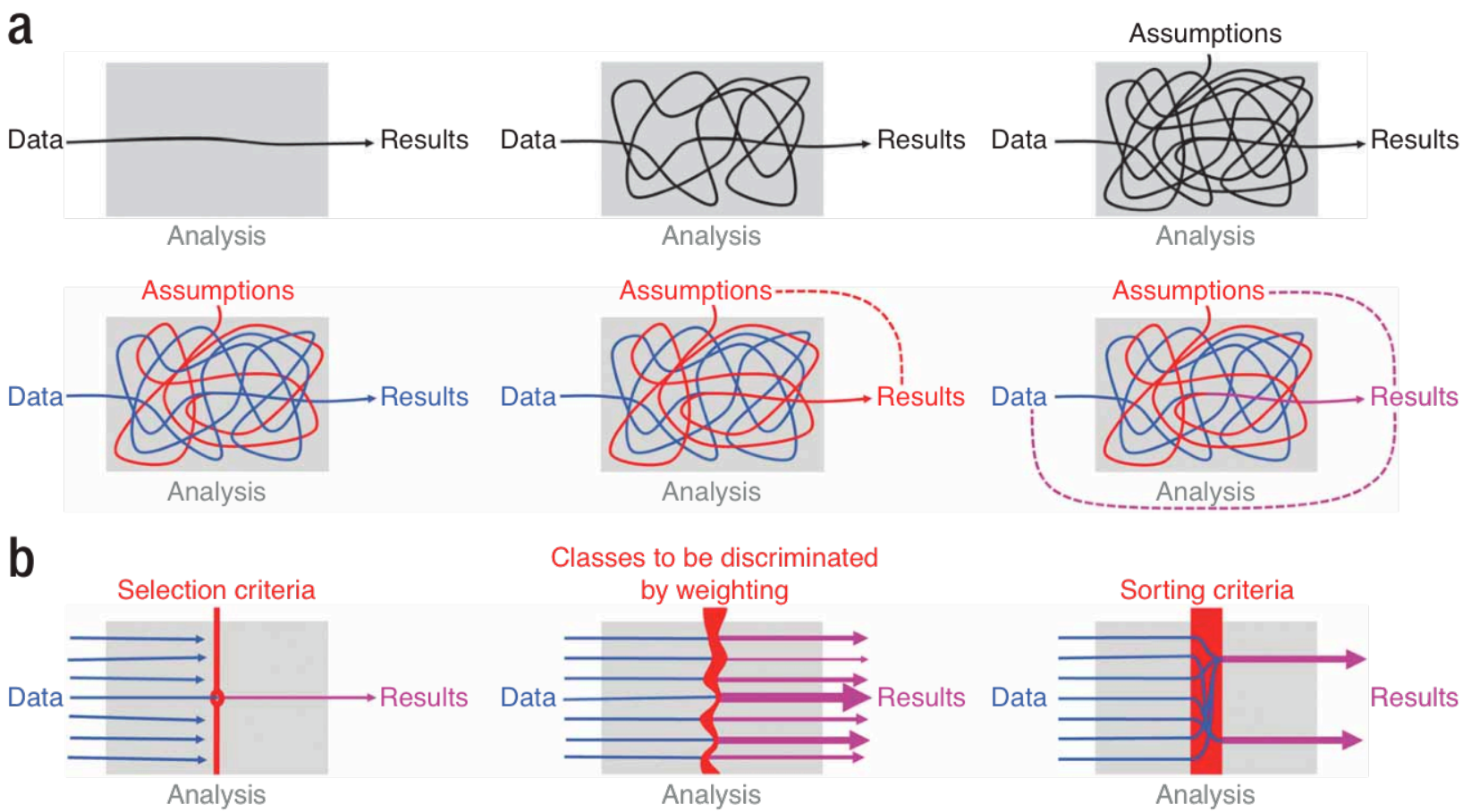
- Bennett et al (209). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Task: open-ended mentalizing versus rest
Stimuli: photographs with valence



minimal cluster size k>8
cluster-level significance p<.001
16 active voxels

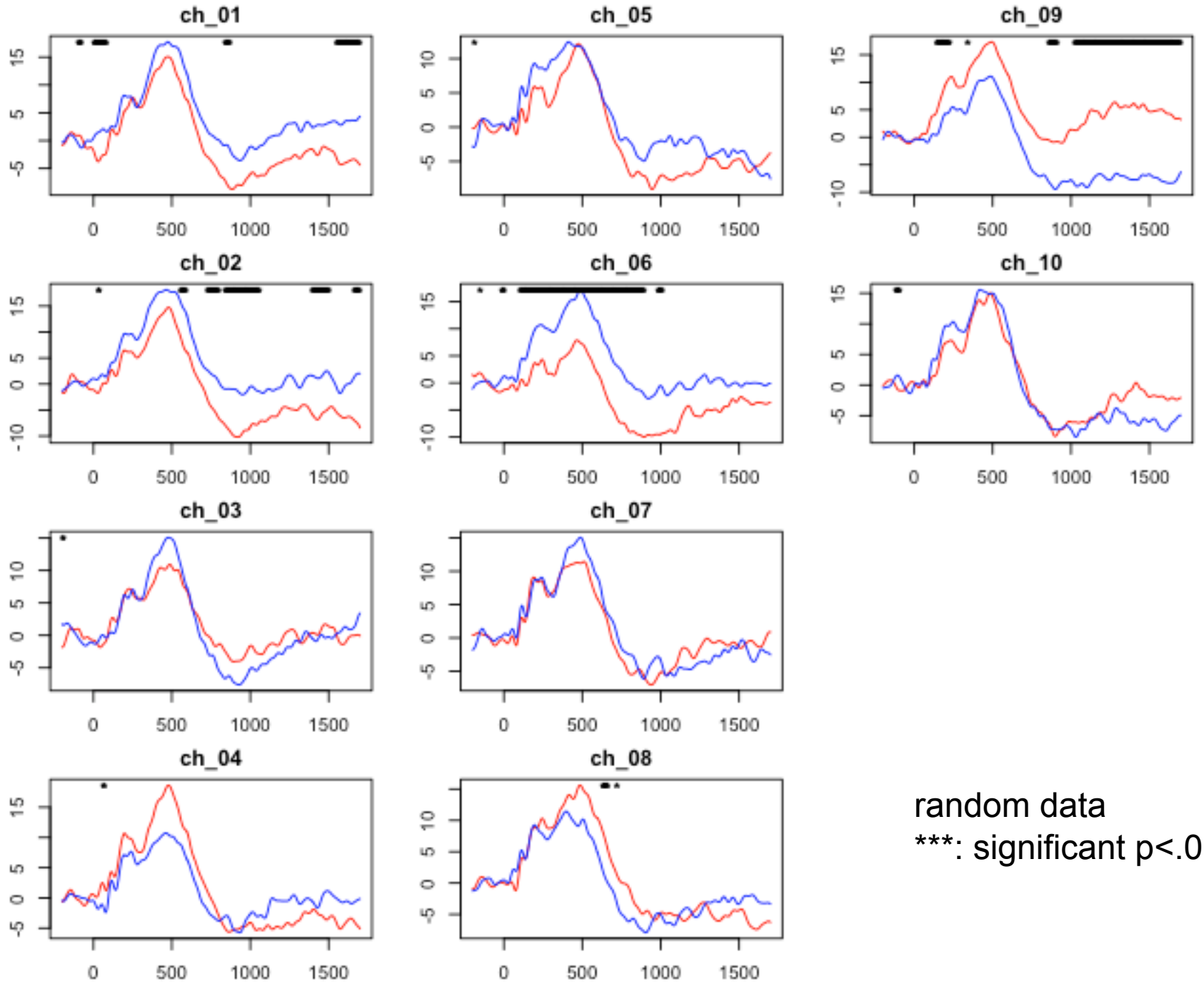
data exploration/double dipping



Kriegeskorte, N. Simmons, W K, Bellgowan P S F & Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. 12, 5, Nat. Neuroscience

Multiple comparison problem

- pervasive in neuroimaging
  - fMRI: voxels
  - EEGs/MEG: time window, frequency window, ROIs
  - all methods: filtering, statistical model, etc



random data
***: significant p<.05

- what to do?
  - fMRI: statistics on voxel cluster size (gaussian fields)
  - general technique: Monte Carlo Resampling of tmax under the null hypothesis

- testing experimenters' degrees of freedom
  - generate a random dataset (eg: 30 random numbers in 3 conditions)
  - select a significance level (eg: .05)
  - construct a test pipeline (eg, a t-test between any of the three conditions)
  - store the results as a series of binary variables (significant, not significant,)
  - redo this N times (eg 10000 times)
  - count the nb of false alarms (it should be 1/20)

- Example of pipelines with degrees of freedom
  - selecting data tranform (eg, three transformation: raw data, log transform, ratio): false alarm rate +50%
  - continuous testing (eg, from 10 to 30 subjects): false alarm rate:+300%
  - condition selection (eg: three conditions): false alarm rate: +280%
  - combination of three three degrees of freedom:

  false alarm rate

  1/20 → 1/2

  to get a real fa rate of 1/20, one needs a p value threshold of .0025, not .05

What to do?

- Supervizor:
  - « look at your data!! »
  - « it is already something if you can make sense of part of your data »
  - « significance thresholds are annoying, what is important is to have good theories »

- BEFORE THE EXPERIMENT
  - specify the experimental plan, the inclusion criterion, region of interest, expected effect sizes, analysis method, nb of subjects
- AFTER THE EXPERIMENT (BUT WITHOUT LOOKING AT THE CONDITIONS)
  - Verifiying the integrity of raw data (outliers, signal/noise, etc).
- PLANNED DATA ANALYSIS
  - no data mining
  - report EVERYTHING
- AFTER PUBLICATION
  - Datamining
  - replication

possible bubble mechanisms

- data collection: experimenter biases
- data analysis: data massaging (this morning)
- write up: biased rendering of the literature, misreporting of data, unconscious plagiarism, authoring issues
- reviewing & publication biases: null results/non replication avoidance, forced citations

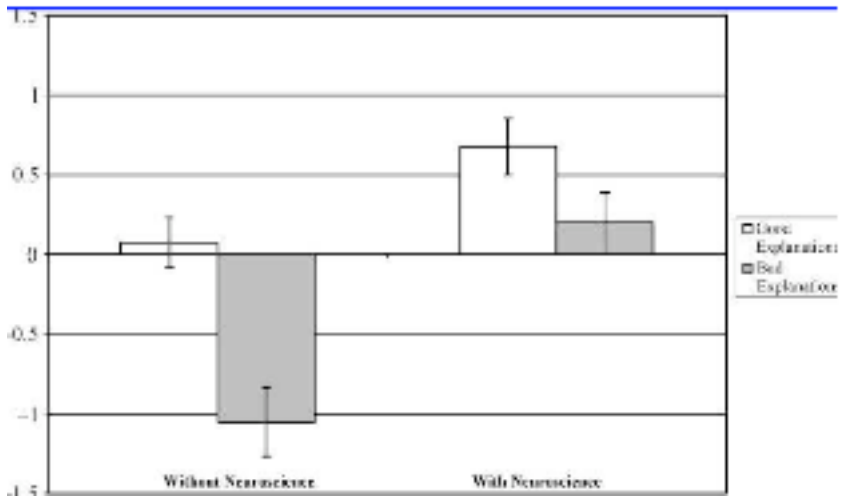

THE SCIENCE NEWS CYCLE
JORGE CHAM © 2009
Start Here
Your Research
Conclusion: A is correlated with B (ρ=0.56), given C, assuming D and under E conditions.
...is translated by...
UNIVERSITY PR OFFICE
(YES, YOU HAVE ONE)
FOR IMMEDIATE RELEASE: SCIENTISTS FIND POTENTIAL LINK BETWEEN A AND B
(UNDER CERTAIN CONDITIONS).
...which is then picked up by...
NEWS WIRE ORGANIZATIONS
A CAUSES B, SAY SCIENTISTS.
...who are read by ...
THE INTERNETS
Scientists out to kill us again.
POSTED BY RANDOM DUDE
Comments (377)
OMG! i kneeew itt!!
WTH???????
...then noticed by...
dingding
We saw it on a Blog!
A causes B all the time
What will this mean for Obama?
BREAKING NEWS BREAKING NEWS BREA
...and caught on ...
CNC Cable NEWS
A=B
4 LOCAL EYEWITLESS NEWS
WHAT YOU DON'T KNOW ABOUT "A"... CAN KILL YOU! MORE AT 11...
A KILLER AMONG US??
LOCAL EYEWITLESS NEWS
...eventually making it to...
I'M WEARING THIS TO WARD OFF "A"
YOUR GRANDMA
WWW.PHDCOMICS.COM

Seducing allure of neuroscience

Researchers created a list of facts that about 50% of people knew. Subjects in this experiment read the list of facts and had to say which ones they knew. They then had to judge what percentage of other people would know those facts. Researchers found that the subjects responded differently about other people's knowledge of a fact when the subjects themselves knew that fact. If the subjects did know a fact, they said that an inaccurately large percentage of others would know it, too. For example, if a subject already knew that Hartford was the capital of Connecticut, that subject might say that 80% of people would know this, even though the correct answer is 50%. The researchers call this finding "the curse of knowledge."

Table 1. Sample Item

	Good Explanation	*Bad Explanation*
Without Neuroscience	The researchers claim that this "curse" happens because subjects have trouble switching their point of view to consider what someone else might know, mistakenly projecting their own knowledge onto others.	The researchers claim that this "curse" happens because subjects make more mistakes when they have to judge the knowledge of others. People are much better at judging what they themselves know.
With Neuroscience	**Brain scans indicate** that this "curse" happens because **of the frontal lobe brain circuitry known to be involved in self-knowledge.** Subjects have trouble switching their point of view to consider what someone else might know, mistakenly projecting their own knowledge onto others.	**Brain scans indicate** that this "curse" happens because **of the frontal lobe brain circuitry known to be involved in self-knowledge.** Subjects make more mistakes when they have to judge the knowledge of others. People are much better at judging what they themselves know.



Weisberg et al (2008). J. Neurosc

what to do with null results?

- publish them!
  - Open Science Collaboration http://openscienceframework.org/
  - Psych File Drawer http://PsychFileDrawer.org/
  - All Results Journals http://www.arjournals.com/ojs/
  - actes de conférence, posters, 'technical reports'
- use/conduct meta-analyses
  - (beware of the publication bias)

- authoring
  - discuss the principles beforehand
  - write down the contributions of each coauthors and aknowledged collaborator
- (see the APA guidelines)
- Plagiarism
- Open Access / Open Science

- <u>Good Practices/gray area: what do to?</u>
    - Différents cas de figure (données, authorship, etc)
  - Q1: Que faire quand on soupçonne/détecte un problème de bonnes pratiques chez nos coauteurs? Tuteurs? collègues?
  - Q2: Quelles recommandations pour la collecte/ stockage des données?
  - Q3: Quelles recommandations pour la lecture et la citation des sources ?

- <u>Good Practices/gray area: what do to?</u>
  - Différents cas de figure (données, authorship, etc)
  - Q1: Que faire quand on soupçonne/détecte un problème de bonnes pratiques chez nos coauteurs? Tuteurs? collègues?
    - Contacter et demander une explication
    - Vérifier les données
    - Informer le reste du groupe
    - Contacter autorités/quelles autorités
  - Q2: Quelles recommandations pour la collecte/stockage des données?
    - Archiver les données brutes dès la collecte, les mettre en libre accès, **utiliser un cahier de laboratoire**
  - Q3: Quelles recommandations pour la lecture et la citation des sources ?
    - Analyser les sources/varier les sources
    - regarder les métaanalyses