

La crise de la recherche non reproductible

- ▶ 10 histoires
- ▶ 10 principes
- ▶ 3 étapes vers une solution

Peut-on utiliser les méthodes bayésiennes pour résoudre la crise des résultats de la recherche statistiquement significatifs que ne tiennent pas?

Andrew Gelman

Département de Statistique et Département des Science Politique,
Columbia University, New York

En visite à ENSAE, Paris

18 mars 2014

Histoire 1: Les attitudes politiques des hommes avec les gros bras (Les problèmes de mesure)

Psychological SCIENCE

A Journal of the
Association for
Psychological Science

[Home](#)[OnlineFirst](#)[All Issues](#)[Subscribe](#)[RSS](#) [Email Alerts](#)

The Ancestral Logic of Politics



Upper-Body Strength Regulates Men's Assertion of Self-Interest Over Economic Redistribution

Over human evolutionary history, upper-body strength has been a major component of fighting ability. Evolutionary models of animal conflict predict that actors with greater fighting ability will more actively attempt to acquire or defend resources than less formidable contestants will. Here, we applied these models to political decision

Histoire 2: La perception extrasensorielle (Les interactions)

Journal's Paper on ESP Expected to Prompt Outrage

By **BENEDICT CAREY**

Published: January 5, 2011

One of psychology's most respected journals has agreed to publish a paper presenting what its author describes as strong evidence for extrasensory perception, the ability to sense future events.

 [Enlarge This Image](#)



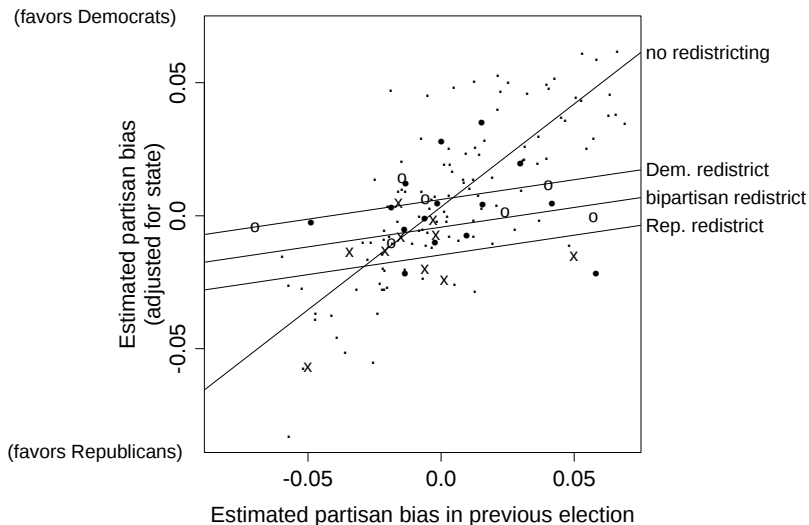
Heather Ainsworth for The New York Times

Work by Daryl J. Bem on extrasensory perception is scheduled to be published this year.

The decision may delight believers in so-called paranormal events, but it is already mortifying scientists. Advance copies of the [paper](#), to be published this year in The Journal of Personality and Social Psychology, have circulated widely among psychological researchers in recent weeks and have generated a mixture of amusement and scorn.

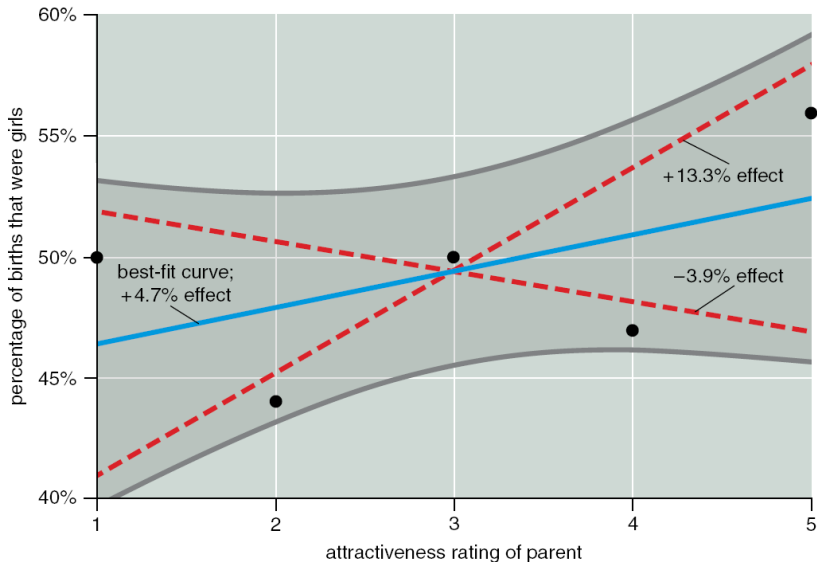
The paper describes nine unusual lab experiments performed over the past decade by its author, [Daryl J. Bem](#), an emeritus professor at Cornell, testing the ability of college students to accurately sense random events,

Histoire 3: Les effets du découpage électoral (Les interactions)



Histoire 4: La beauté et les proportions des sexes

(Les estimations qui sont trop grandes d'être plausibles)



Histoire 5: L'ovulation et la couleur des vêtements

(Degrés de liberté de chercheur)

Psychol Sci. 2013 Sep 1;24(9):1837-41. doi: 10.1177/0956797613476045. Epub 2013 Jul 10.

Women are more likely to wear red or pink at peak fertility.

Beall AT, Tracy JL.

University of British Columbia.

Abstract

Although females of many species closely related to humans signal their fertile window in a certain manner, often involving red or pink coloration, no such display has been found for humans. To provide evidence that men are sexually attracted to women wearing or surrounded by red, we tested whether women show a behavioral tendency toward wearing reddish clothing when at peak fertility. In two samples (N = 124), women at high conception risk were more than 3 times more likely to wear a pink shirt than were women at low conception risk, and 77% of women who wore red or pink were found to be at high, rather than low, risk. Conception risk had no effect on the prevalence of wearing red or pink color. Our results thus suggest that red and pink adornment in women is reliably associated with peak fertility and that female ovulation, long assumed to be hidden, is associated with a salient color display.

Histoire 6: L'ovulation et la vote

(Les estimations qui sont trop grandes d'être plausibles)

The Fluctuating Female Vote: Politics, Religion, and the Ovulatory Cycle

**Kristina M. Durante¹, Ashley Rae¹, and
Vladas Griskevicius²**

¹College of Business, University of Texas, San Antonio, and ²Carlson School of Management, University of Minnesota

Abstract

Each month, many women experience an ovulatory cycle that regulates fertility. Although the cycle influences women's mating preferences, we proposed that it might also change women's political views. Building on theory suggesting that political and religious orientation are linked to reproductive goals, we tested how fertility influenced women's politics, religiosity, and voting in the 2012 U.S. presidential election. Results showed that women with large and diverse families had relatively different effects on political orientation and voting behavior compared to women with small families.

Histoire 7: La subjectivité et les singes (Les problèmes de mesure)

Marc Hauser Resigns From Harvard



By Tom Bartlett

Marc D. Hauser, the Harvard psychologist found responsible for eight counts of misconduct by the university, has ended speculation about whether the embattled professor would return to Harvard this fall.

In a [letter](#) dated July 7, Mr. Hauser told Michael D. Smith, Harvard's dean of the

Histoire 8: La recherche inventé (La fraude)

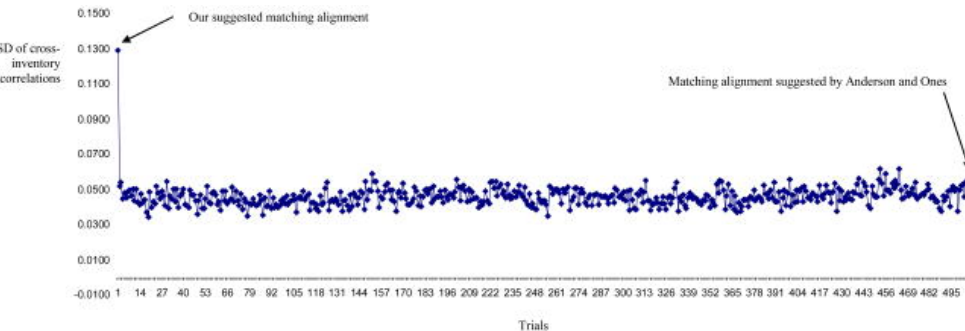
 [Enlarge This Image](#)



Joris Buijs/Pve

The psychologist, Diederik Stapel, committed academic fraud in “several” papers, many accepted in respect in the news media, according to a report Monday by the three Dutch institutions that worked: the University of Groningen, the University of Amsterdam, and Tilburg. The journal published one of Dr. Stapel’s papers as an “editorial expression of concern” online on Tuesday.

Histoire 9: «Aucune preuve irréfutable» (Les erreurs du traitement des données)



Histoire 10: L'intervention précoce chez l'enfant (Peu de données)

Charles Murray: “To me, the experience of early childhood intervention programs follows the familiar, discouraging pattern . . . small-scale experimental efforts [$N = 123$ and $N = 111$] staffed by highly motivated people show effects. When they are subject to well-designed large-scale replications, those promising signs attenuate and often evaporate altogether.”

James Heckman: “The effects reported for the programs I discuss survive batteries of rigorous testing procedures. They are conducted by independent analysts who did not perform or design the original experiments. The fact that samples are small works *against* finding any effects for the programs, much less the statistically significant and substantial effects that have been found.

October 22, 2013

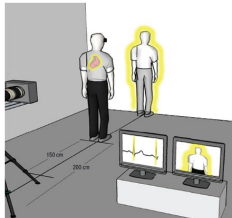


This Week in *Psychological Science* (TWiPS)

The links below take you to the journal via the APS website. If not already logged in, you will be redirected to log-in using your last name (Gelman) and Member ID (8167).

[Turning Body and Self Inside Out: Visualized Heartbeats Alter Bodily Self-Consciousness and Tactile Perception](#)

Jane Elizabeth Aspell, Lukas Heydrich, Guillaume Marillier, Tom Lavanchy, Bruno Herbelin, and Olaf Blanke



Studies of body perception have mostly focused on manipulations of exteroceptive cues (e.g., vision and touch); however, interoceptive cues (i.e., representations of internal bodily states) may be just as important for self-perception. Participants viewed a virtual body or a rectangle, each of which had a flashing outline that was synchronous or asynchronous with the participant's own heartbeat. Self-identification was stronger for people viewing the virtual body with the synchronous flashing outline than for those viewing the body with the asynchronous flashing outline or for those viewing the rectangles. This suggests that both interoceptive and exteroceptive cues play

important roles in bodily self-perception.

[Aging 5 Years in 5 Minutes: The Effect of Taking a Memory Test on Older Adults' Subjective Age](#)

Matthew L. Hughes, Lisa Geraci, and Ross L. De Forrest

Subjective age – how old people feel – is related to psychological and physical well-being. In this study, the researchers examined whether common memory-testing procedures influence adults' subjective age. Older and younger adults rated their subjective age before and after taking a memory test. Older adults reported feeling older after taking the memory test, but younger adults did not. A follow-up study found that simply anticipating taking a memory test increased older adults' subjective age. These



Cette semaine dans *Psychological Science*

- ▶ “Turning Body and Self Inside Out: Visualized Heartbeats Alter Bodily Self-Consciousness and Tactile Perception”
- ▶ “Aging 5 Years in 5 Minutes: The Effect of Taking a Memory Test on Older Adults’ Subjective Age”
- ▶ “The Double-Edged Sword of Grandiose Narcissism: Implications for Successful and Unsuccessful Leadership Among U.S. Presidents”
- ▶ “On the Nature and Nurture of Intelligence and Specific Cognitive Abilities: The More Heritable, the More Culture Dependent”
- ▶ “Beauty at the Ballot Box: Disease Threats Predict Preferences for Physically Attractive Leaders”
- ▶ “Shaping Attention With Reward: Effects of Reward on Space- and Object-Based Selection”
- ▶ “It Pays to Be Herr Kaiser: Germans With Noble-Sounding Surnames More Often Work as Managers Than as Employees”

Cette semaine dans *Psychological Science*

- ▶ $N = 17$
- ▶ $N = 57$
- ▶ $N = 42$
- ▶ $N = 7\,582$
- ▶ $N = 123 + 156 + 66$
- ▶ $N = 47$
- ▶ $N = 222\,924$

Principe 1: La difference entre «significatif» et «non significatif» n'est pas elle-même statistiquement significative

- ▶ Expérience 1: 25 ± 10 : significatif
- ▶ Expérience 2: 10 ± 10 : bruit!
- ▶ Différence: 15 ± 14 , 1: ...

Principe 2: Un loi a priori non informatif donne les inférences qu'on ne doit pas croire

- ▶ Expériment 2: 10 ± 10 : bruit!
- ▶ Mais, avec la loi a priori non informatif, $\Pr(\text{effet réel} > 0) = 0.84!$
- ▶ Les études épidémiologiques avec 95% intervalle de confiance $[1, 1, 8.5]$

Principe 3: Les hypothèses de recherche et les hypothèses de statistique

Dans un cas, on veut confirmer; dans l'autre, on doit rejeter.

- ▶ La perception extrasensorielle
- ▶ Les gros bras

Principe 4: Le filtre de la signification statistique

Les estimations statistiquement significatives sont au moyen trop grandes.

- ▶ La beauté et les proportions des sexes
- ▶ L'intervention précoce chez l'enfant

Principe 5: Degrés de liberté de chercheur

C'est pas seulement l'effet tiroir.

- ▶ Les gros bras
- ▶ Le découpage électoral

Principe 6: Le Jardin aux sentiers qui bifurquent

Les degrés de liberté de chercheur peut être un problème même si il n'y a pas «la pêche» des valeurs p et même si l'hypothèse de recherche était annoncé à l'avance

- ▶ Degrés de liberté de chercheur
- ▶ L'ovulation et la vote

Principe 7: L'erreur, «Ce qui ne détruit pas ma signification statistique la rend plus forte»

C'est une intuition qui ne marche pas quand la variation est grande

- ▶ L'ovulation et la couleur des vêtements
- ▶ L'intervention précoce chez l'enfant

Principe 8: La poursuite de la certitude

Quelles sont les similarités entre la pratique de la recherche habituelle, les problèmes éthiques, et la fraude?

- ▶ La desire pour la certitude
- ▶ Les motivations pour apparaître certain
- ▶ L'attitude de ne jamais revenir vers le bas

- ▶ Les exemples de *Psychological Science*
- ▶ La perception extrasensorielle
- ▶ Les chercheurs qui disait, «Aucune preuve irréfutable»
- ▶ Hauser et Stapel

Principe 9: Les erreurs de type S et type M

- ▶ J'ai jamais fait un «error de type 1» dans ma vie
- ▶ J'ai jamais fait un «error de type 2» dans ma vie
- ▶ Je fais les erreurs de type S (signe)
- ▶ Je fais les erreurs de type M (magnitude)

Principe 10: La variation and les interactions

Les interactions sont important et sûrement existent mais sont difficiles d'estimer avec la bonne précision

- ▶ La subjectivité et les singes
- ▶ Les exemples de *Psychological Science*

- ▶ Les données publiques (inclus les protocoles de mesure, les formulaires d'enquête, et les informations sur le traitement et l'analyse des données)
- ▶ Publier des études réussies ou non
- ▶ Publication visible de rétractations, critiques, et répétitions
- ▶ Répétition avec la préinscription

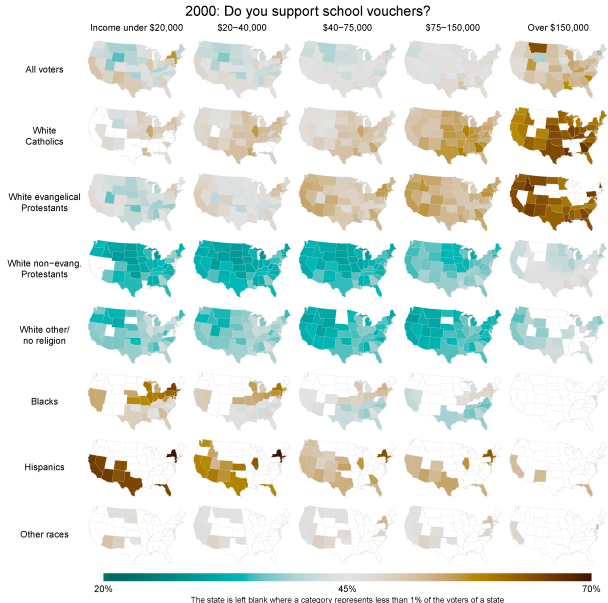
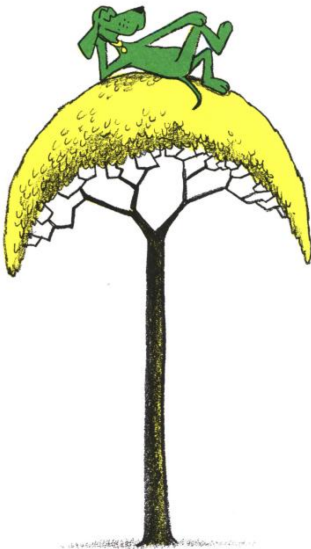
Solution 1: Les calculations pour dessin d'analyse

- ▶ Une généralisation du concept d'analyse de puissance statistique
- ▶ Estimate cru: les parents plus beaux sont 4,7% plus probable d'avoir les filles (avec erreur type 4,3%)
- ▶ Suppose la différence vraie est 0,3%
- ▶ La calculations pour dessin d'analyse
 - ▶ 3% probabilité d'un résultat statistiquement significative et positif
 - ▶ 2% probabilité d'un résultat statistiquement significative et *negatif*
 - ▶ Le taux des erreurs de type S est 40%
 - ▶ Le facteur d'hausse de type M est au moins $\frac{1.96*4.3\%}{0.3\%} = 28$

Solution 2: Le loi a priori informatif

- ▶ On peut le réaliser par Bayes ou par une calcul pour dessin d'analyse
- ▶ Les proportions des sexes: on est presque sûr que la différence réelle est dans l'intervalle $(-0.3\%, +0.3\%)$
- ▶ Ovulation et la vote: une différence dans l'intervalle $(-2\%, 2\%)$
- ▶ Ici, 3 types de la croyance a priori:
 - ▶ La différence est presque 0 (la plupart des idées ne marchent pas, diminution a cause des erreurs de mesure, etc)
 - ▶ La différence est positif (la croyance du chercheur)
 - ▶ La différence est négatif (le biais vers le pessimisme)

Solution 3: Les modèles hiérarchiques pour les interactions



Les données ne parlent pas toujours d'eux-mêmes

