Open Access as an antidote for the self-referential character of science

Alessandro Giuliani

Istituto Superiore di Sanita’, Roma

That's incredible ! You scientists are the only category that pays for writing an article instead of being paid (my dad, some years ago)

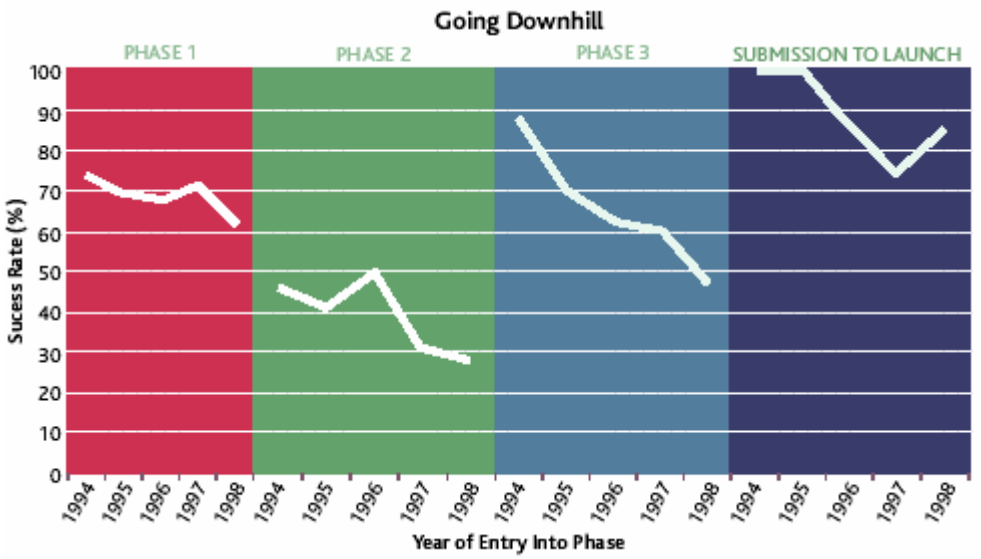
The above paradox is still more emphasized by Open Access,
The writers continue to pay, the readers not.

Scientific papers have the same status of advertising.

If scientific literature can be equated to advertising,
the second question is:

What are the goods do scientists advertise ?

The declining trend of new drugs

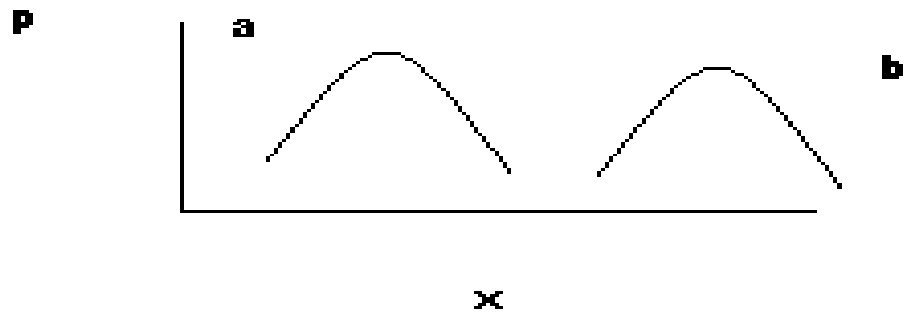


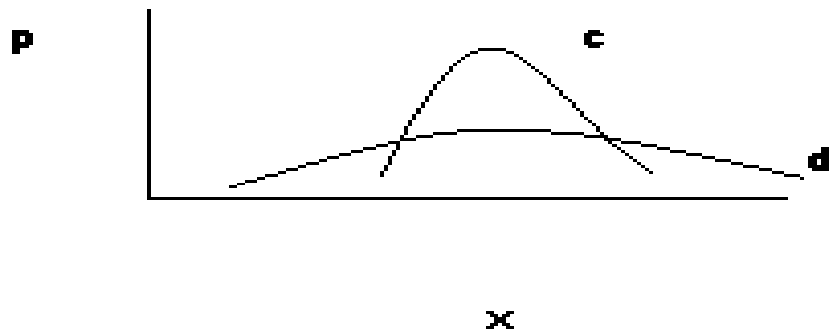
And what about pure knowledge ?

Come and have a look to the fabric where it is built….

$E(X) = \Sigma (X(i))/N)$

Std. Dev. $(X) = \sqrt{\Sigma (X(i)) - E(X))^2 / N}$

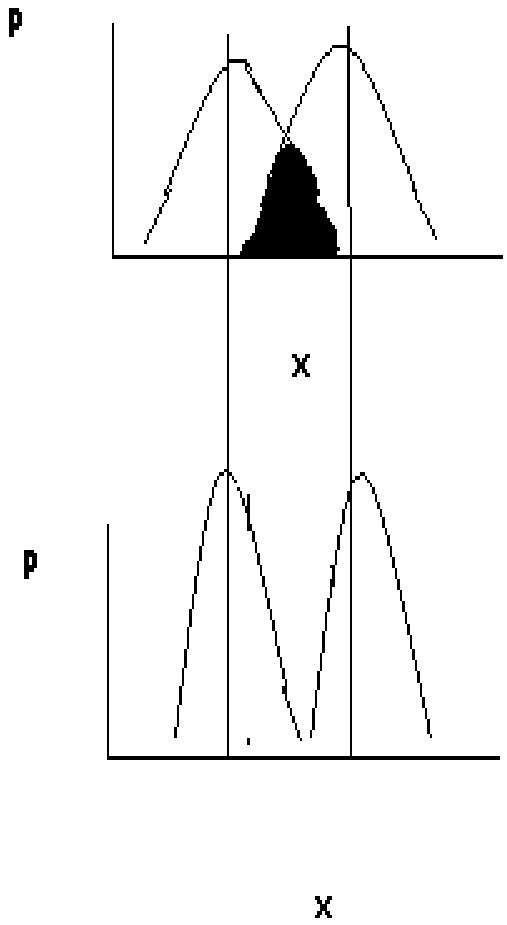




Going from descriptive to inferential statistics we shift our interest from the Natural Variability (i.e. how much an individual differs from its population mean) to our approximation in knowing the ‘real value of the population mean’.

This is a completely different problem, by the way the population mean could have no meaning at all.

In any case, while natural variability (SD) is a measure having a characteristic scale, our approximation on the population mean can be made as little as we want by increasing our sampling.



$$ES = \text{Std.Dev.} / \sqrt{N} \qquad (3)$$

The magic formula

$$-1.96*ES < (M1 - M2) < 1.96*ES$$

=

?

$$ES = Std.Dev. / \sqrt{N}$$

Open access, freely available online

Essay

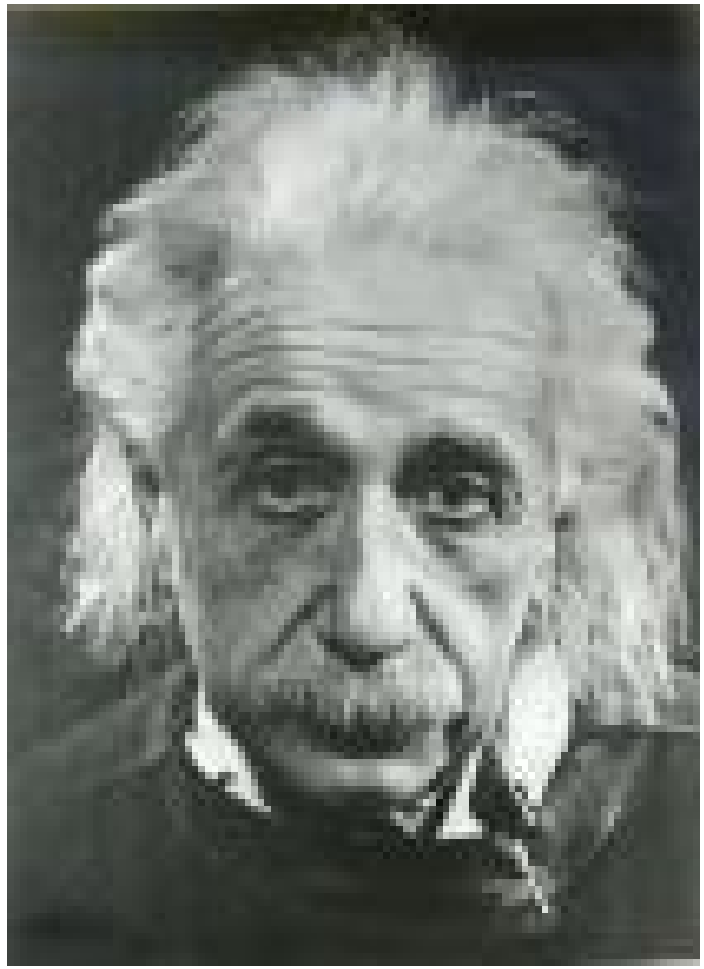
Why Most Published Research Findings Are False

John P. A. Ioannidis

August 2005 | Volume 2 | Issue 8 | e124

Table 4. PPV of Research Findings for Various Combinations of Power (1 − β), Ratio of True to Not-True Relationships *(R)*, and Bias *(u)*

1 − β	*R*	*u*	Practical Example	PPV
0.80	1:1	0.10	Adequately powered RCT with little bias and 1:1 pre-study odds	0.85
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41
0.20	1:5	0.20	Underpowered, but well-performed phase I/II RCT	0.23
0.20	1:5	0.80	Underpowered, poorly performed phase I/II RCT	0.17
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015



http://www.dis.uniroma1.it/~farina/semplice/

28–31 | PNAS | January 4, 2000 | vol. 97 | no. 1

The Theory of Everything

R. B. Laughlin* and David Pines†‡§

*Department of Physics, Stanford University, Stanford, CA 94305; †Institute for Complex Adaptive Matter, University of California Office of the President, Oakland, CA 94607; ‡Science and Technology Center for Superconductivity, University of Illinois, Urbana, IL 61801; and §Los Alamos Neutron Science Center Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Contributed by David Pines, November 18, 1999

We discuss recent developments in our understanding of matter, broadly construed, and their implications for contemporary research in fundamental physics.

The Theory of Everything is a term for the ultimate theory of the universe—a set of equations capable of describing all phenomena that have been observed, or that will ever be observed (1). It is the modern incarnation of the reductionist ideal of the ancient Greeks, an approach to the natural world that has been fabulously successful in bettering the lot of mankind and continues in many people's minds to be the central paradigm of physics. A special case of this idea, and also a beautiful instance of it, is the equation of conventional nonrelativistic quantum mechanics, which describes the everyday world of human beings—air, water, rocks, fire, people, and so forth. The details of this equation are less important than the fact that it can be written down simply and is completely specified by a handful of known quantities: the charge and mass of the electron, the charges and masses of the atomic nuclei, and Planck's constant. For experts we write

we have learned why atoms have the size they do, why chemical bonds have the length and strength they do, why solid matter has the elastic properties it does, why some things are transparent while others reflect or absorb light (6). With a little more experimental input for guidance it is even possible to predict atomic conformations of small molecules, simple chemical reaction rates, structural phase transitions, ferromagnetism, and sometimes even superconducting transition temperatures (7). But the schemes for approximating are not first-principles deductions but are rather art keyed to experiment, and thus tend to be the least reliable precisely when reliability is most needed, i.e., when experimental information is scarce, the physical behavior has no precedent, and the key questions have not yet been identified. There are many notorious failures of alleged *ab initio* computation methods, including the phase diagram of liquid ^{3}He and the entire phenomenonology of high-temperature superconductors (8–10). Predicting protein functionality or the behavior of the human brain from these equations is patently absurd. So the triumph of the reductionism of the Greeks is a pyrrhic victory: We have succeeded in reducing all of ordinary

Biomedical Digital Libraries

BioMed Central

Review

Open Access

Conceptual biology, hypothesis discovery, and text mining: Swanson's legacy

Tanja Bekhuis*

Address: Department of Library & Information Science, School of Information Sciences, University of Pittsburgh, 135 North Bellefield Avenue, Pittsburgh, PA 15260, USA. Current address: Department of Biology, Juniata College, 1700 Moore Street, Huntingdon, PA 16652, USA

Email: Tanja Bekhuis* - bekhuis@juniata.edu

* Corresponding author

Published: 03 April 2006

*Biomedical Digital Libraries*2006, 3:2 doi:10.1186/1742-5581-3-2

Received: 26 September 2005

Accepted: 03 April 2006

This article is available from: http://www.bio-diglib.com/content/3/1/2

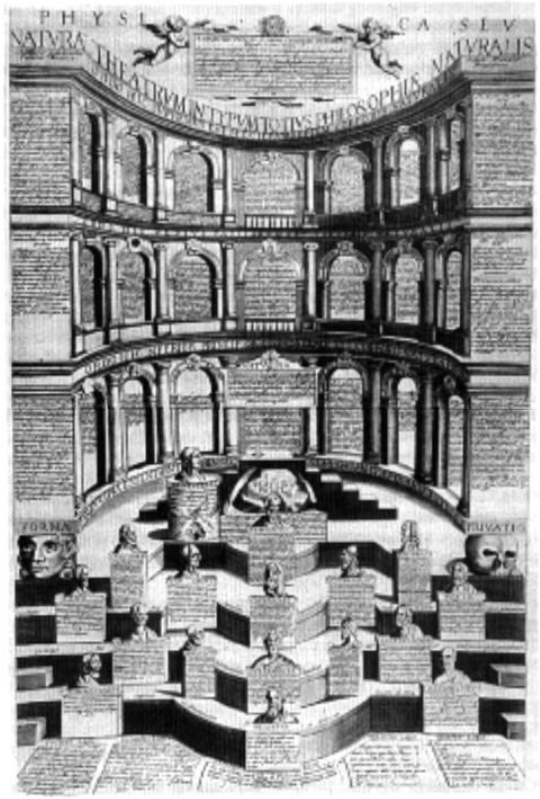
VNIVERSÆ
NATVRAE
THEATRVM.

IN QVO RERVM OMNIVM
effectrices causæ, & fines contemplantur, &
continua series quinque libris
discutiuntur.

AVTORE
IOAN. BODINO.

FRANCOFVRTI,
Apud heredes Andreæ Wecheli,
Claudium Marnium, & Ioan. Aubr.
M. D. XCVII

Jean Bodin [1597] *Universae naturae theatrum,*



Philander Colutius [1611]
Theatrum naturae,

Microparadigms: Chains of collective reasoning in publications about molecular interactions

Andrey Rzhetsky*†‡§, Ivan Iossifov*†, Ji Meng Loh¶, and Kevin P. White‖

*Department of Biomedical Informatics, †Columbia Genome Center, and ‡Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032; ¶Department of Statistics, Columbia University, New York, NY 10027; and ‖Department of Genetics, Yale University, New Haven, CT 06520

Communicated by Sherman M. Weissman, Yale University School of Medicine, New Haven, CT, January 23, 2006 (received for review August 15, 2005)

We analyzed a very large set of molecular interactions that had been derived automatically from biological texts. We found that published statements, regardless of their verity, tend to interfere with interpretation of the subsequent experiments and, therefore, can act as scientific "microparadigms," similar to dominant scientific theories [Kuhn, T. S. (1996) *The Structure of Scientific Revolutions* (Univ. Chicago Press, Chicago)]. Using statistical tools, we measured the strength of the influence of a single published statement on subsequent interpretations. We call these measured values the momentums of the published statements and treat separately the majority and minority of conflicting statements about the same molecular event. Our results indicate that, when building biological models based on published experimental data, we may have to treat the data as highly dependent-ordered sequences of statements (i.e., chains of collective reasoning) rather than unordered and independent experimental observations. Furthermore, our computations indicate that our data set can be interpreted in two very different ways (two "alternative universes"): one is an "optimists' universe" with a very low incidence of false results (<5%), and another is a "pessimists' universe" with an extraordinarily high rate of false results (>90%). Our computations deem highly unlikely any milder intermediate explanation between these two extremes.

Bayesian inference | quality of science | text mining | experiment interpretation | information cascade

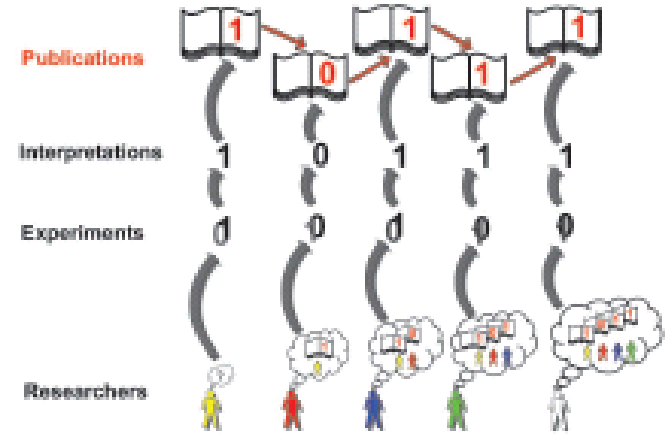


Fig. 1. A hypothetical chain of collective reasoning. The chain is started by a scientist who performs an experiment hidden from the outside world. The results of the experiment involve some fuzziness, and the chain originator publishes the most likely interpretation given the absence of prior publications. The second, third, and all other scientists who join the chain later, think in the context of the published opinions and can be led to interpret their experimental results differently than would be done in the absence of prior data. The fourth and fifth persons in the chain publish interpretations of their data that would be opposite in the absence of prior publication.

Scientists are becoming more and more a closed clique having only endogamic relations, this is extremely dangerous for the progress of their work.

Music has already experienced this kind of effect, Mozart was listened by the average man of his times, only other dodecaphonic composers can listen to dodecaphonic music.

A person of average culture must understand the real thing not the divulgative version of science.

This is a crucial point in these times

IDENTIFICAZIONE DEI GRUPPI OMOGENEI

Per segmentare le imprese oggetto dell'analisi in gruppi omogenei sulla base degli aspetti strutturali, si è ritenuta appropriata una strategia di analisi che combina due tecniche statistiche:

- una tecnica basata su un approccio di tipo multivariato, che si è configurata come un'analisi fattoriale del tipo *Analyse des données* e nella fattispecie come un'*Analisi in Componenti Principali* [1];
- un procedimento di *Cluster Analysis* [2].

L'utilizzo combinato delle due tecniche è preferibile rispetto a un'applicazione diretta delle tecniche di clustering.

In effetti, tanto maggiore è il numero di variabili su cui effettuare il procedimento di classificazione, tanto più complessa e meno precisa risulta l'operazione di clustering.

Per limitare l'impatto di tale problematica, la classificazione dei contribuenti è stata effettuata a partire dai risultati dell'analisi fattoriale, basandosi quindi su un numero ridotto di variabili (i fattori) che consentono, comunque, di mantenere il massimo delle informazioni originarie.

In un procedimento di clustering di tipo multidimensionale, quale quello adottato, l'omogeneità dei gruppi deve essere interpretata, non tanto in rapporto alle caratteristiche delle singole variabili, quanto in funzione delle principali interrelazioni esistenti tra le variabili esaminate che contraddistinguono il gruppo stesso e che concorrono a definirne il profilo.

Le variabili prese in esame nell'Analisi in Componenti Principali sono quelle presenti in tutti i quadri di cui si compone il questionario ad eccezione del quadro M che contiene gli stessi dati contabili presenti nella dichiarazione dei redditi. Tale scelta nasce dall'esigenza di caratterizzare le imprese in base ai possibili modelli organizzativi, alle diverse tipologie di prodotti venduti, alle varie modalità di acquisto, etc.; tale caratterizzazione è possibile solo utilizzando le informazioni relative alle strutture operative, al mercato di riferimento e a tutti quegli elementi specifici che caratterizzano le diverse realtà economiche e produttive di una impresa.

I fattori risultanti dall'Analisi in Componenti Principali vengono analizzati in termini di significatività sia economica sia statistica, al fine di individuare quelli che colgono i diversi aspetti strutturali delle attività oggetto dello studio.

La Cluster Analysis ha consentito di identificare sette gruppi omogenei di imprese.

Open Access constitutes a first step in letting some air to flow into the closed world of scientists: even people outside the research centers can in principle read first-hand science

The second step must be done by scientists forcing themselves to be simple by describing only the relevant portion of their work

This second step is surely possible.
I had personally this experience.
Multidimensional statistics is the main tool for this goal to be achieved

Even if you cannot play basket as
Michael Jordan plays,
you can understand he is a champion

The same can be done with science