Commentary

A contemporary pathology of science

Carlo Modonesi[a], Lorenzo Farina[b], Ignazio Licata[c], Roberto Germano[d], Joseph. P. Zbilut[e] and Alessandro Giuliani[f]

[a] *Museo di Storia Naturale, Università degli Studi, Parma, Italy*
[b] *Dipartimento di Informatica e Sistemistica, Università degli Studi "La Sapienza", Rome, Italy*
[c] *Institute for Scientific Methodology, Palermo, Italy*
[d] *Promete inc., Naples, Italy*
[e] *Department of Molecular Biophysics and Physiology, Rush University Chicago, USA*
[f] *Dipartimento di Ambiente e Connessa Prevenzione Primaria, Istituto Superiore di Sanità, Rome, Italy*

Summary. A contemporary pathology of science is outlined. This pathology suggests that "previous knowledge" drastically limits innovative thinking in science. In very raw "Bayesian" terms it is affirmed that a too rich and flexible *a priori* knowledge is detrimental to the appreciation of novelty coming from experimental results by both lowering the relative weight assigned to *a posteriori* contrasting evidence and adapting potentially revolutionary findings to an already existing frame.

Key words: epistemology, statistics, physics, pharmacology, computational biology.

INTRODUCTION: AN HISTORICAL PREMISE

At the dawn of modern science, in the period from the end of the 16th century to the first half of the 17th, the intellectual class of Europe was convinced that the basic scientific understanding of the natural world was already achieved, and that the important task facing the brilliant minds of the time was the organization of knowledge into general schemata they called *Theatrum Mundi* (Theater of the World) or *Theatrum Naturae* (Theater of Nature). There were many of these "general systems" theories, the most famous of which are probably those by Kircher and Fludd [1, 2].

In the same epoch, the brilliant scientist Johannes Kepler, was addressing the problem of the description of the orbit of Mars. The classical way to accomplish this task was by means of the so called epicycle calculus. This kind of computation enabled astronomers to obtain perfect representations of planet trajectories. As in the spirit of *Theatrum Mundi*, the trick was generating a paradigm (in this case, a circular orbit) that could be applied to any kind of phenomenon to obtain a perfect description by simple iteration. We now know why epicycles worked so well: they are nothing more than a different version of the Fourier spectrum of a trajectory, which is supported by a theorem that says that every signal, no matter how complex, can be traced back to a convolution of sine waves.

Kepler, however, made a fundamental move out of the sterile perfection of epicycle calculus, and thus provided an important impetus to the birth of modern science: he assumed orbits were not circular but elliptic. This approach simplified computations, did not worsen (neither ameliorate) the quality of predictions, and in essence, was simply an aesthetic move toward a more "beautiful" explanation, which initiated the opportunity for driving astronomy out of the sterile repetition of "adequate" descriptions.

The pathology of epicycles has a name: overfitting, which more or less means that if you have too many adjustable parameters to explain a given phenomenon, you will arrive at a perfect fitting of the phenomenon itself. The simple reason for this is that you are not discovering anything new but simply rephrasing the problem.

Clinical medicine too, during its development, was not immune from these "all-encompassing / impossible-to-falsify" theories. Three examples (among many others) are bloodletting, bed rest and miasmas theories.

Address for correspondence: Alessandro Giuliani, Dipartimento di Ambiente e Connessa Prevenzione Primaria, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy. E-mail: alessandro.giuliani@iss.it.

Bloodletting was for centuries considered an important therapy for diverse diseases. Associated with the idea of "humors" and their excess, it was used up until the 20th century. As late as 1901 Sir William Osler [3] used it for, among other diseases, pneumonia. Although still used in specific cases (*e.g.*, polycythemia), advances in pathology and treatment have discredited its routine use.

Although its history is not clearly established, the idea that bed rest is an important part of therapy for illness has persisted until the present day, at least in the popular mind. This is despite research demonstrating the deleterious effects of prolonged immobility such as clots, hypotension, and bone demineralization [4].

The theory of "miasmas" developed in the Middle Ages in conjunction with plagues. It identified the causative agent of diseases such as cholera as emanating from the poisonous particles of decomposition. During this time, physicians wore masks filled with sweet smelling flowers. Although clearly wrong, it did stimulate hygienic practices which did, in fact, improve the public health. Gradually the germ theory took over; nonetheless, it persisted to the 19th century, and is related to the Italian word "malaria", *i.e.*, "bad air".

What makes these theories a kind of "medical epicycle"? First, their extreme flexibility: all three paradigms can be adapted to practically any situation and context with very minor modifications. Secondly, the fact they have a "hidden tautological explanation"; *i.e.*, they are identified as the causative agents for any success (like epicycle with Fourier analysis). In the case of bloodletting, it was not uncommon for the patient to be fed with food more nutritious than usual, thus effecting a general improvement of his natural defenses (consider the fact that many of these patients were malnourished). Similar reasoning holds for bed rest, which was a pause to very stressful conditions and physical fatigue, while for miasmas the "Fourier analysis" was the presence of different pathogens like *Plasmodium falciparum* carried by mosquitos in swamps smelling of decomposing biological matter.

These contingencies made these theories practically not falsifiable. We are still in presence of these kinds of theories in contemporary medicine, and the practice of meta-analysis can be thought as a kind of response to this problem.

Our point is that nowadays science is passing through an epicycle-like pathology and in the rest of the paper we will try to illustrate at least some of the features of this pathology how they appear in different scientific fields.

THE HEAVY BURDEN OF PREVIOUS KNOWLEDGE

Molecular medicine and biology are especially prone to "theories of everything". Currently, databases of genes (so called ontologies), proteins, interactions, and pathways coming from various "omics" of biomedical sciences are the contemporary version of "whole-world" representations of four centuries ago. Bioinformatic methods allow for the vitiation of virtually any observed experimental result by the use of already stored information that is so rich and flexible that it is practically impossible to be falsified.

By mining previous literature or electronic repositories, we can find support for any hypotheses. We have sufficiently knowledge to explain everything or nothing. This paradoxical (and extremely dangerous for the advancement of science) situation is very well described in the paper by Rzhetsky *et al.* [5] who explicitly calls "microparadigms", ideas that, regardless of their verity, become more and more pervasive in the scientific community because the results they come from are no longer under scrutiny, and their consequences are simply taken as a an undoubted truth to build other models. The advancement in information technology, progressively increasing the automation of data base compilation, worsens the situation, especially in the case of data (such as protein interaction experiments) with a heavy dynamical character. However, the development of new biochemical techniques has falsified much of the already existing interaction data, and the risk of building castles on unreliable experimental data is extremely high.

In the meantime, notwithstanding the huge amount of stored information, molecular pharmacology seems to become more and more ineffective: the number of new drugs on the market has rapidly fallen over the last three decades. Overington and colleagues [6] estimated that 76% of new drugs developed from 1989 to 2003 refer to targets already known before modern molecular biology, and only 6% bind to new targets. The so-called biotechnology drug revolution does not seem to live up to initial expectations.

The same regressive tendencies described for biomedical sciences can be found in theoretical physics as well. They are displayed in the a-critical adherence to "Everything-or-nothing theories". Such theories consider fundamental problems as Relativity, Quantum theory, and the "thermodynamic balloon" of the "Big Bang" in several versions, which, still contain the potential for unsolved problems. The difficulty with theories of almost-everything is that the "almost" which they do not describe, (*e.g.*, the open problem in Quantum

Field Theory and Stochastic Quantization), is more interesting than the "everything" they do describe [7].

Again scientists seem to follow the mythology of the "complete picture" from which they derive all phenomena by purely formal means: theoretical physics becomes more and more self-referential, dangerously reaching a *Theatrum Mundi* of its own with no possibility of explaining any observed natural phenomena.

The possibility to give, in perfect good faith, a plausible explanation to any accidental finding by a pure re-arrangement of a too large and unstructured knowledge *corpus* has recognized effects on the loss of efficiency of science. This mechanism is at the basis of the so called "Protheus effect" mentioned by Ioannidis [8] in which he recognizes a vital cycle common to a lot of scientific "breakthroughs" starting with some brilliant data confirmed by different research groups and ending with a progressive recognition of the illusory character of the finding. These life cycles tend to be shorter and shorter with the passage of time.

We strongly believe the presence of an unprecedented massive burden of previous data (theories in the case of physics, vastly accepted evidences in biology and medicine) to be "honoured" and "acknowledged" by new discoveries mitigate the possibility of innovation from new experiments. This "pathology" opens the way for a purely rhetorical use of scientific arguments; while at the same time severely hampers the possibility for new paradigms to come in play.

We need more quality, not quantity, in producing scientific knowledge. This is to say we need new heuristic approaches in research practice, as well as new social education and participation in governing its contributions to culture and well being.

CONCLUSIONS

We love science, and this is the reason why we write this brief essay, and why we feel the need to point out some critical elements for planning new policies of science and innovation. We think that true innovation suffers from a too tight embrace with a too big and too flexible corpus of previous knowledge. We must dare to re-start with new paradigms, new data, new methods and use previous knowledge only as "statistical material" to be submitted to experimentation and not as established golden standards.

We must dare to be brutal and refuse to listen to "super experts". We must dare to collect new data with new criteria (*e.g.*, there is an embarrassing paucity of reliable and sufficiently long time series data for metabolism, gene expression, protein abundance, etc.). We must favor model-free and easy to test exploratory approaches over more sophisticated ones, which, however, tend to be very dependent in terms of unjustified assumptions (*e.g.*, complex parametric models in which an overwhelming number of parameters are fitted to small data sets). We must dare to start scientific enterprises without the pressing need to be immediately "translational" otherwise we will be confined to the already known, already accepted, and probably without a future.

References

1. Kircher A. *Polygraphia nova et universalis ex combinatoria arte detecta.* Rome: Varesius; 1663.
2. Fludd R. *Utriusque cosmi historia.* Oppenheim; 1617-19. Available from: http://www.billheidrick.com/Orpd/RFludd/index.htm; last visited 2/09/2008.
3. Osler W. *Principles and practice of medicine.* New York: Appleton; 1901.
4. Allen C, Glasziou P, Del Mar C. Bed rest: a potentially harmful treatment needing more careful evaluation. *Lancet* 1999;354(9186):1229-33.
5. Rzhetsky A, Iossifov I, Meng Loh J, White KP. Microparadigms: Chains of collective reasoning in publications about molecular interactions. *Proc Natl Acad Sci USA* 2006;103:4940-5.
6. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov* 2006;5:993-6.
7. Woit P. Not Even Wrong: *The failure of string theory and the search for unity in physical law.* New York: Basic Books; 2006.
8. Ioannidis J. Why most published research findings are false. *PLoS Medicine* 2005;2(8):e124.