Bad Science in the Age of Petabytes

Tony Tyson

Physics Department

UC Davis

Responsible Conduct of Research Seminar
December 6, 2007

The Scientific Process

Systematic Error happens

So does Human Nature

Skepticism vs Enthusiasm

The mark of Pathology

The road to Fraud

Uninformed Analysis

Look at your Data!

But what if you cannot?

The Tyranny of Terabytes

Petabytes: A New World

Automated Discovery, Cyber Analytics

Scientific process

(1) Observation. Experiment design is very important.

(2) Compare with theory. The inducement of general hypotheses or possible explanations for what has been observed. *The simplest hypothesis is the best*. A viable theory must be falsifiable.

(3) The deduction of corollary predictions that must be true if the hypothesis is true. Additional testable predictions are made, based on the initial hypothesis.

(4) Testing the hypothesis by investigating and confirming the deduced implications.

Real discoveries of phenomena contrary to all previous scientific experience are very rare, while fraud, fakery, foolishness, and error resulting from over-enthusiasm and delusion are all too common.

A. Cromer 1993

systematics

Particularly troubling today is that we don't fully know what we don't know

Testimony by Bert Ely to the Subcommittee on Financial Management, the Budget, and International Security of the Senate Committee on Governmental Affairs July 21, 2003

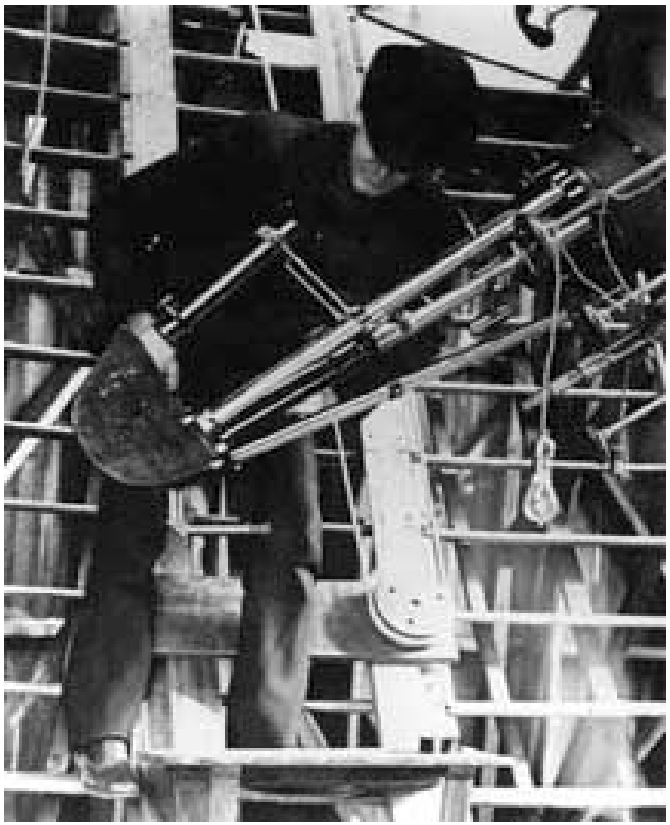
Two kinds of error:

Random error

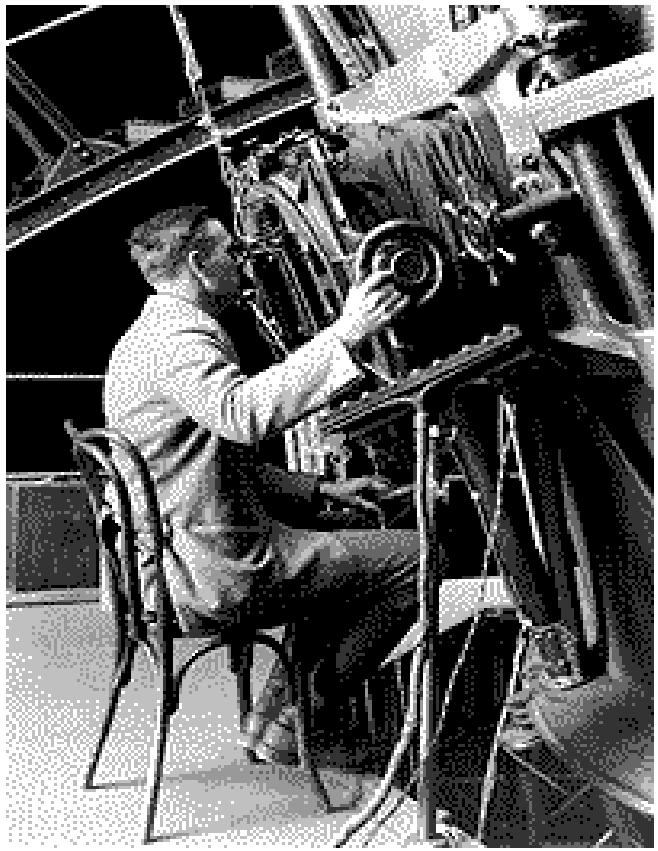
Systematic error

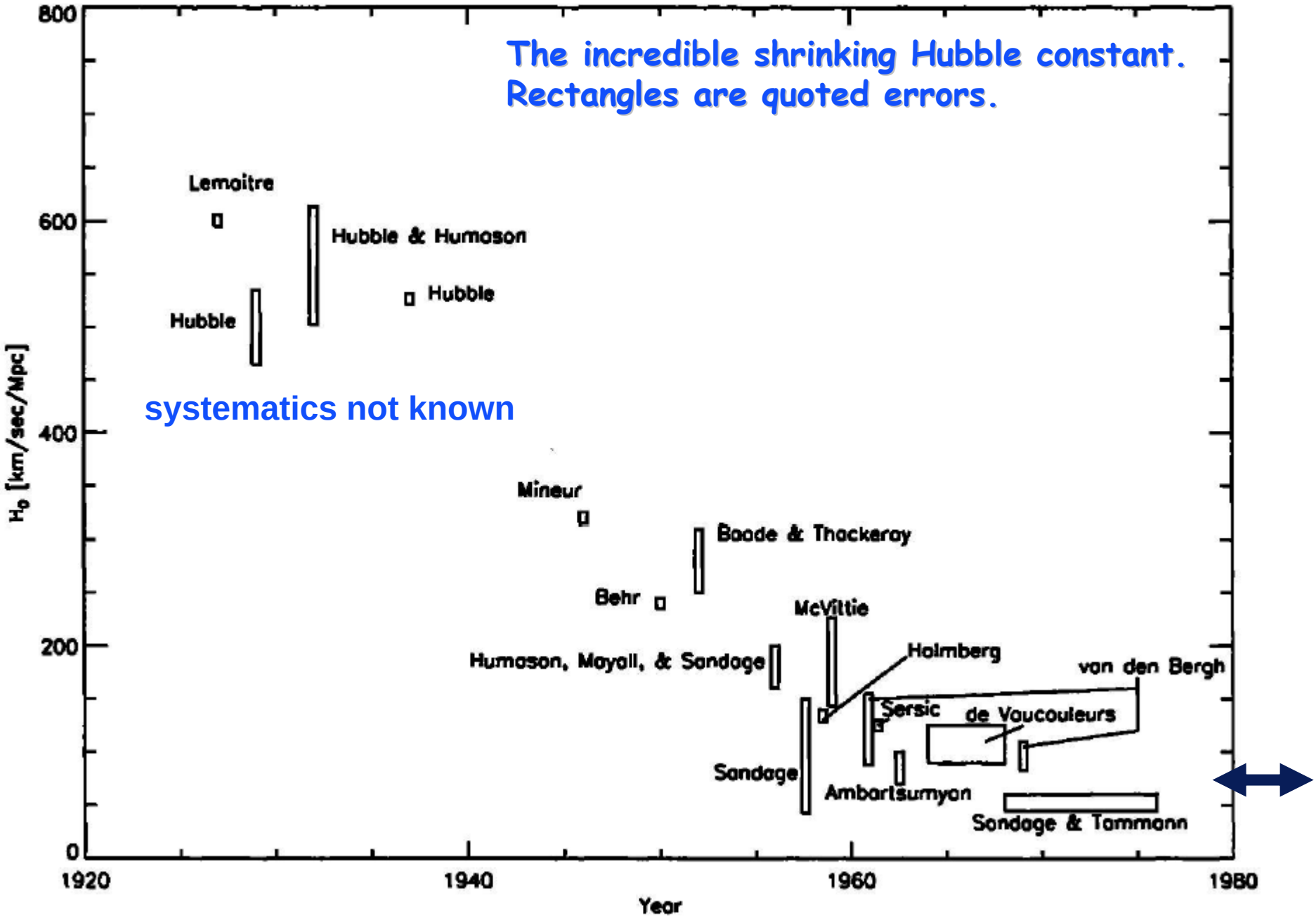
Discovery of expanding universe

Vesto Slipher



Edwin Hubble





Trimble (1996) PASP 108, 1073

Systematics: catch-22

The difficulty is this: if we understand the systematic we can correct for it, but if we don't understand the systematic we won't think of it at all or our error estimate will be wrong.

It is only at the edge of understanding where systematic errors are meaningful: we understand enough to realize it might be a problem, but not enough to easily fix it.

Avoiding Systematics

The best prevention of systematic error is good experiment design.

How can we robustly attack this problem in an existing experiment or observation?

A mix of simulations and exploratory tests.

Simulations are useful teachers of where sensitivity to systematics are. We may then explore these avenues; search for the signature of each systematic, isolate it, understand it, and gain control of it. In practice, for each experimental field it is a kind of "art" which demands familiarity with the likely systematics. It is the responsibility of the experimentalist to probe for systematics and of the theorist to allow for them.

Healthy skepticism

- **Be skeptical of your own work**
- **Test relentlessly for systematics**
- **Avoid early press conferences**

Pathological science

Not fraud

Well intentioned, enthusiastic scientists are led astray

Examples abound in every field of science

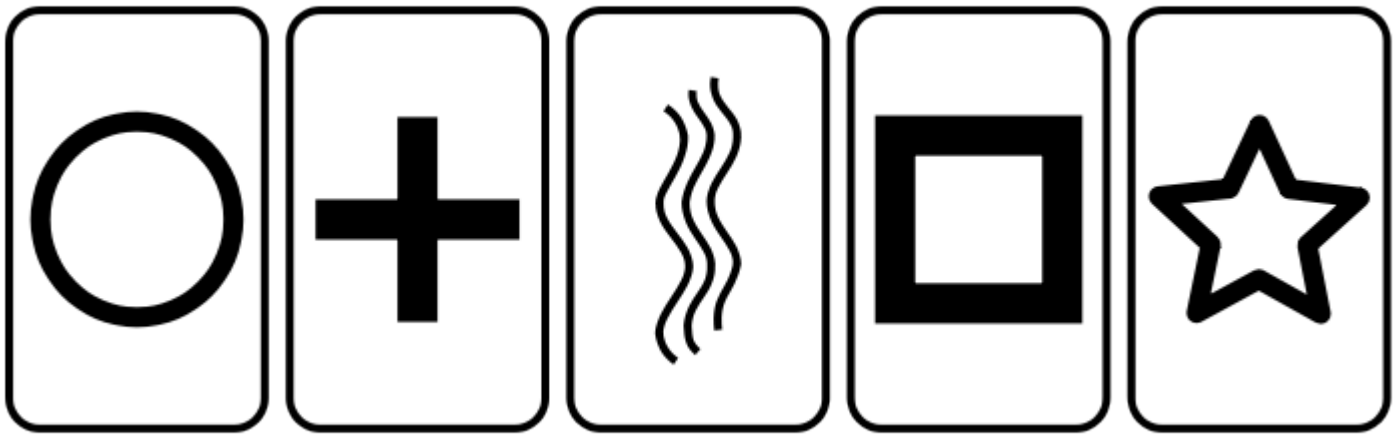
polywater

The case of polywater demonstrates how the desire to believe in a new phenomenon can sometimes overpower the demand for solid, well-controlled evidence. In 1966 the Soviet scientist Boris Derjaguin lectured in England on a new form of water that he claimed had been discovered by another Soviet scientist, N. N. Fedyakin. Formed by heating water and letting it condense in quartz capillaries, this "anomalous water," had a density higher than normal water, a viscosity 15 times that of normal water, a boiling point higher than 100 degrees Centigrade, and a freezing point lower than zero degrees. Over the next several years, hundreds of papers appeared in the scientific literature describing the properties of what soon came to be known as polywater. Theorists developed models, supported by some experimental measurements, in which strong hydrogen bonds were causing water to polymerize. Some even warned that if polywater escaped from the laboratory, it could autocatalytically polymerize all of the world's water.

Then the case for polywater began to crumble. Because polywater could only be formed in minuscule capillaries, very little was available for analysis. When small samples were analyzed, polywater proved to be contaminated with a variety of other substances, from silicon to phospholipids. Electron microscopy revealed that polywater actually consisted of finely divided particulate matter suspended in ordinary water. Gradually, the scientists who had described the properties of polywater admitted that it did not exist. They had been misled by poorly controlled experiments and problems with experimental procedures. As the problems were resolved and experiments gained better controls, evidence for the existence of polywater disappeared.

Extrasensory perception

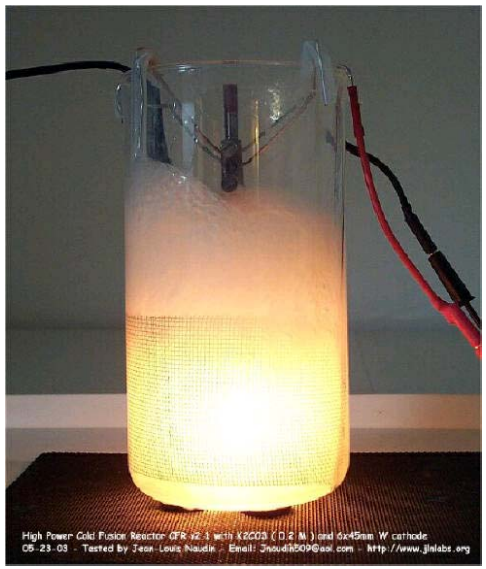
- **Parapsychologist J. B. Rhine (1934)**
- **Common systematic error in “paranormal” statistics**



“Remote viewing” experiments showing a significant effect have one thing in common: only one judge, the principal investigator, was used in all the remote-viewing experiments.

Cold fusion

- Pons and Fleischman claimed bench-top fusion using a palladium battery
- Before peer review, they held a press conference



“Cold fusion” has since been debunked.

Features of Pathological Science

- **The maximum effect is produced by a barely perceptible cause, and the effect doesn't change much as you change the magnitude of the cause.**

- **The effect only happens sometimes, when conditions are just right, and no one ever figures out how to make it happen reliably. The people who can do it are unable to communicate how they make it happen to the people who can't.**

- **The effect is always close to the limit of detectability.**

- **There are claims of great accuracy, well beyond the state of the art or what one might expect.**

- **Fantastic theories contrary to experience are suggested. Often, mechanisms are suggested that appear no where else.**

- **Criticisms are met by ad hoc excuses thought up on the spur of the moment.**

Irving Langmuir 1953 see: *Physics Today* Oct. 1989

Is it pathological?

A single hit does not mark an idea as pathological, but multiple hits should serve to raise one's suspicions. This is a list primarily aimed at experiments, but many of the characteristics can also apply to theories.

Good science can often have one or two of these symptoms. This is because most experiments at the frontier deal with barely detectable signals.

There is always risk in undertaking such experiments (or interpreting them). But there is also great *opportunity!*

Related sociology

- Supporters are unable or unwilling to think about testing or disproving the effect. Tests that could lead to definitive disproof are never done by supporters.

- The implications of a theory or experiment are never extended outside its original domain. Supporters don't ask what implications it might have for neighboring fields.

- The ratio of supporters to critics rises rapidly to ~50% and then slowly decays to zero over a long time.

Good reading

Robert L. Park. *Voodoo Science*: *The Road from Foolishness to Fraud.* Oxford University Press, New York, 2000. ISBN: 0-19-513515-6.

Rousseau, Denis L. *Case Studies in Pathological Science*. American Scientist 80: 54-63 (1992)

Some common mistakes

Poor experiment design

Not testing for systematics (control)

Ignoring sample selection effects (bias)

Bad statistics: assume wrong distribution (tails!)

Failure to repeat the experiment using different sample with same physics

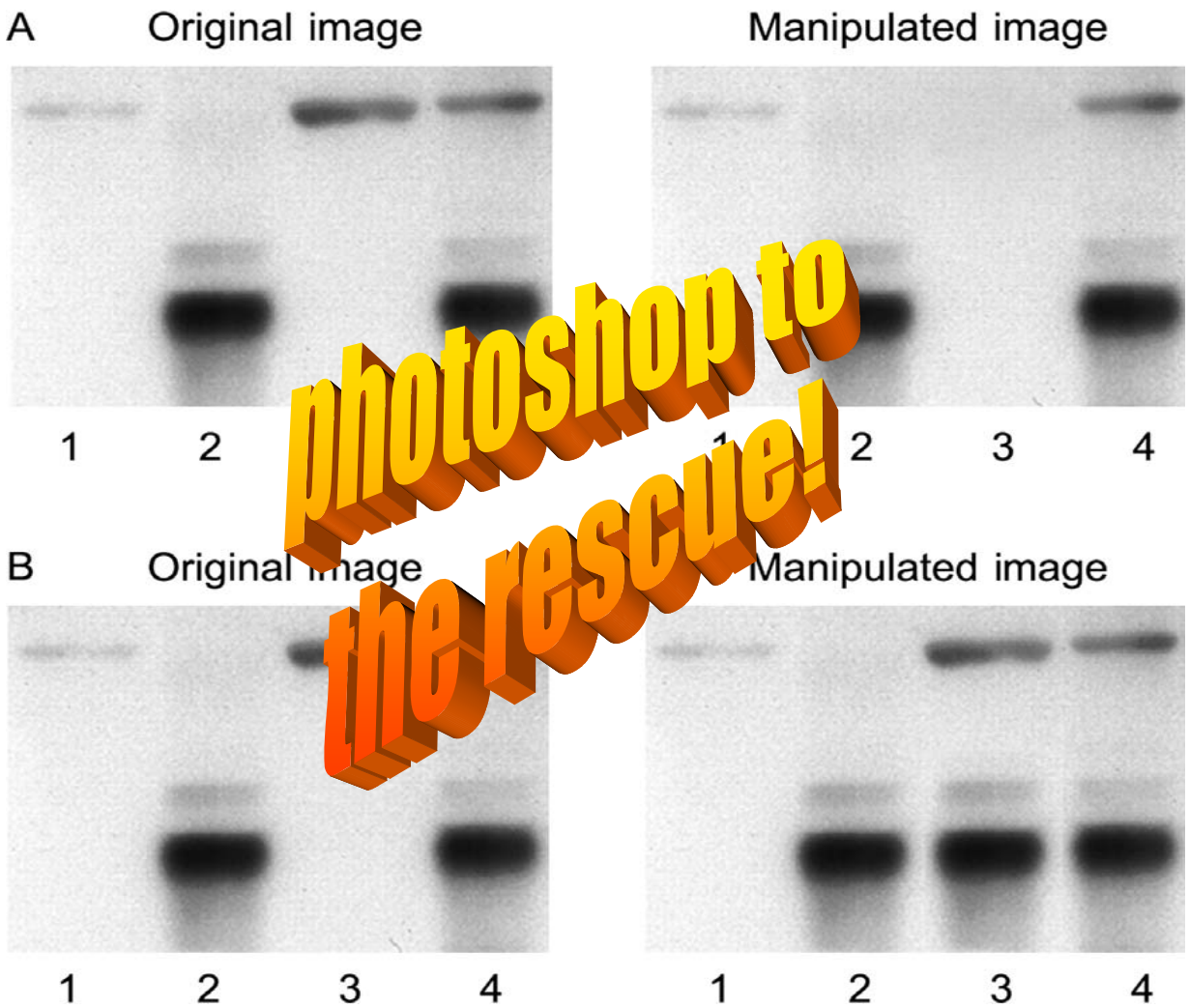
Image Manipulation

Journals find that authors have manipulated data in order to "enhance" appearance and support th

Sometimes th es the line between standard acce d image processing (to remove artifacts of the detector) and fraud.

PUBLISHED IMAGE DATA: MEGABYTES NOW SOON GIGABYTES AND THEN TERABYTES

Figure 1. Gross manipulation of blots



Rossner, M. et al. J. Cell Biol. 2004;166:11-15

Jan Henedrik Schön

Schön joined Bell Labs in 1998, just before finishing his Ph.D. in Konstanz, Germany. In February 2000, Schön published some startling experimental results.

Schön and his partners had started with molecules that don't ordinarily conduct electricity, and claimed they had succeeded in making them behave like semiconductors. The researchers reported their findings in Science.

Less than five years after finishing graduate school, Jan Hendrik Schön was in contention for the Nobel prize.

Schön and collaborators

For two and a half years, Jan Hendrik Schon of Bell Labs was the poster boy for productivity in physics research. During that time, Schon was the lead author of 89 papers. This practically superhuman pace averages to a paper every 10 days.

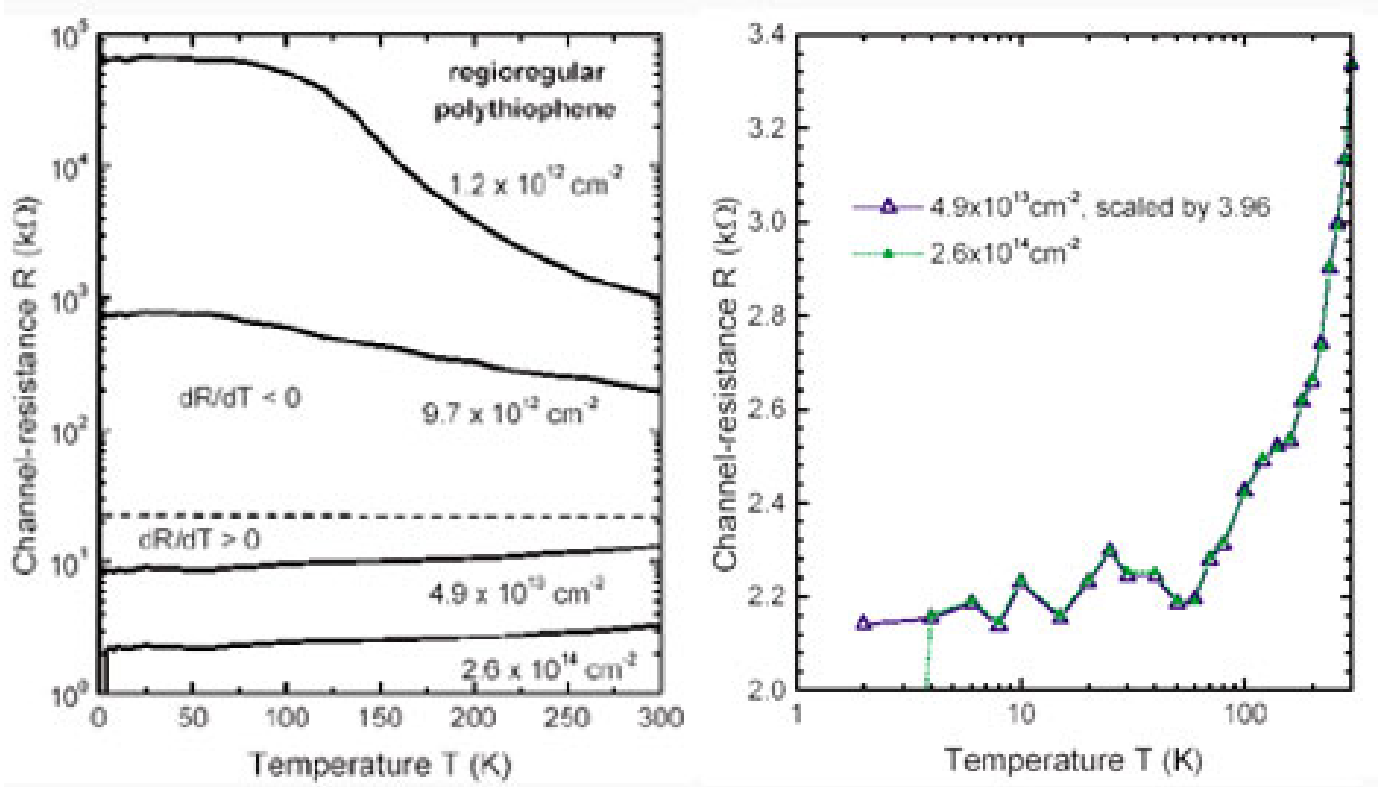
The unraveling

Concern arose when some researchers within Bell Labs told Lydia Sohn of Princeton University that they had noticed a strong resemblance between two papers by three of their Bell Labs colleagues, one appearing in *Nature* and the other in *Science*. Both papers described field-effect transistors made from self-assembled monolayers of organic materials, but the two papers dealt with slightly different materials.

One morning she got in to work to find a message waiting on her answer phone. Prof LYDIA SOHN: "I just happened to check my voicemail messages in my office and I had a very interesting voicemail message and it said, Lydia this is your homework, look at these two papers by Hendrik. And by the tone of his voice I knew something very juicy was going on and so I quickly downloaded the, these two papers, one from Science and one from Nature. "

BBC interview, 2004

Data substitution



Data substitution was found in a paper describing gate-induced superconductivity in polythiophene. The published figure (left panel) shows resistance for four values of surface charge density. Superconductivity sets in at the highest density. The bottom two curves are replotted in the right panel, with one curve divided by 3.96. An investigation found that the data were the same, except for one point.

Science magazine on the Beasley committee report

We have been asked whether this sad incident has given us doubts about how well the peer review process at *Science* works. Unhappy experiences should generate efforts to learn from them, and we will use the report to evaluate what we might have done differently in these cases. That said, we would reiterate that it is asking too much of peer review to expect it to immunize us against clever fraud.

Take risks

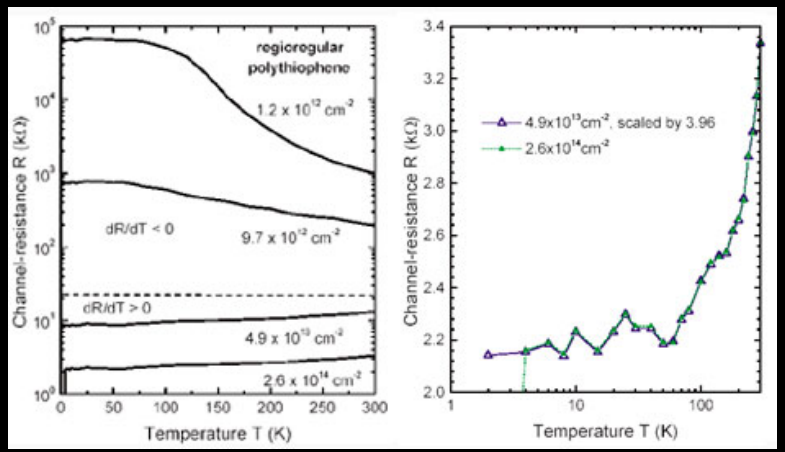
Exploration and discovery involves risk-taking

Interplay between theory and observation (experiment)

Take care

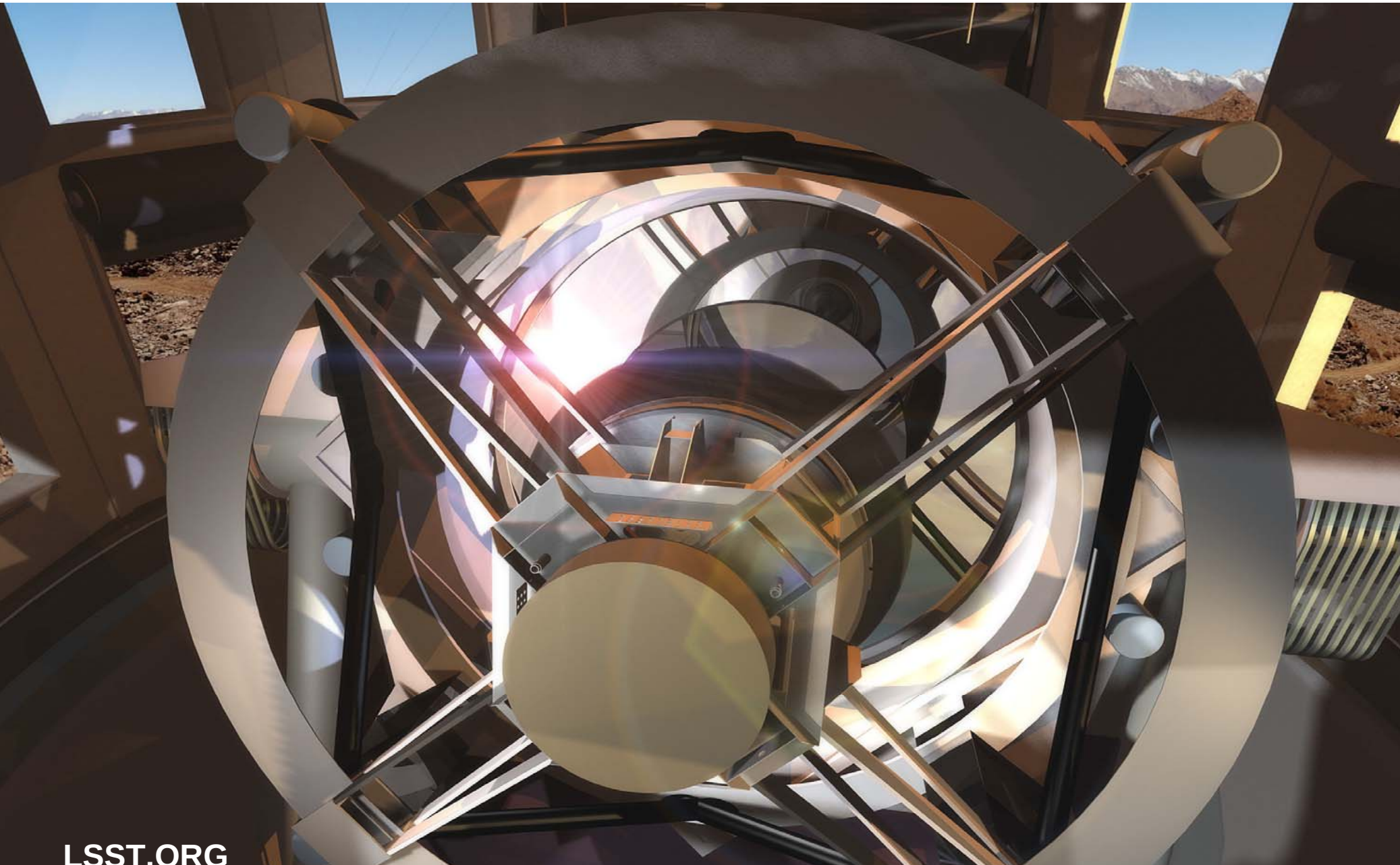
The scientific process demands integrity

Recall how Schon's research was debunked:
Interested scientists LOOKED at his data.
i.e. they used the brain's pattern recognition
to uncover unexpected correlations.



What if there is SO MUCH DATA that the human mind cannot begin to search for patterns?

The New Sky: Celestial Cinematography



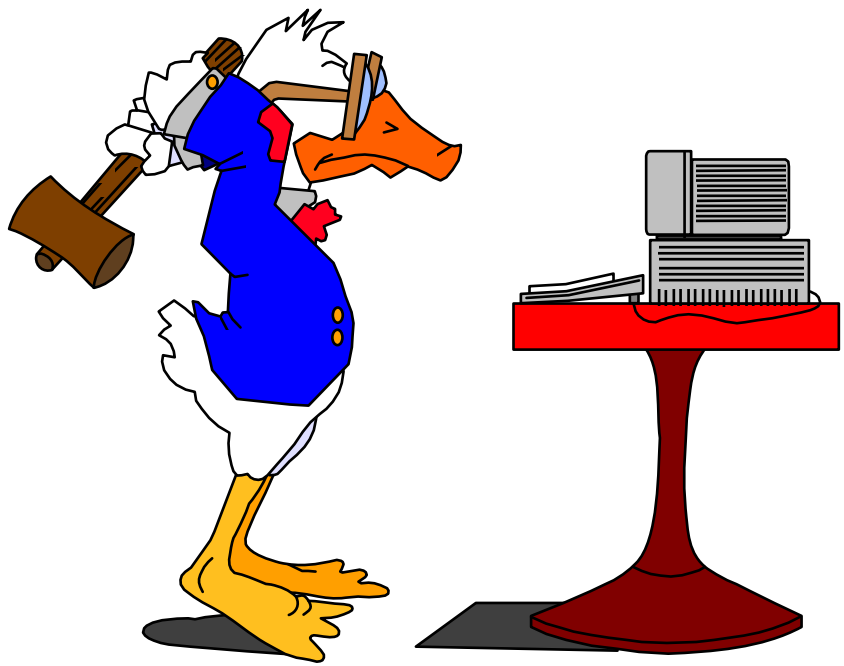
The LSST Data Challenge

- **~2 Terabytes per hour that must be mined in real time.**
- **More than 10 billion objects will be monitored for important variations in real time.**
- **Knowledge extraction in real time.**

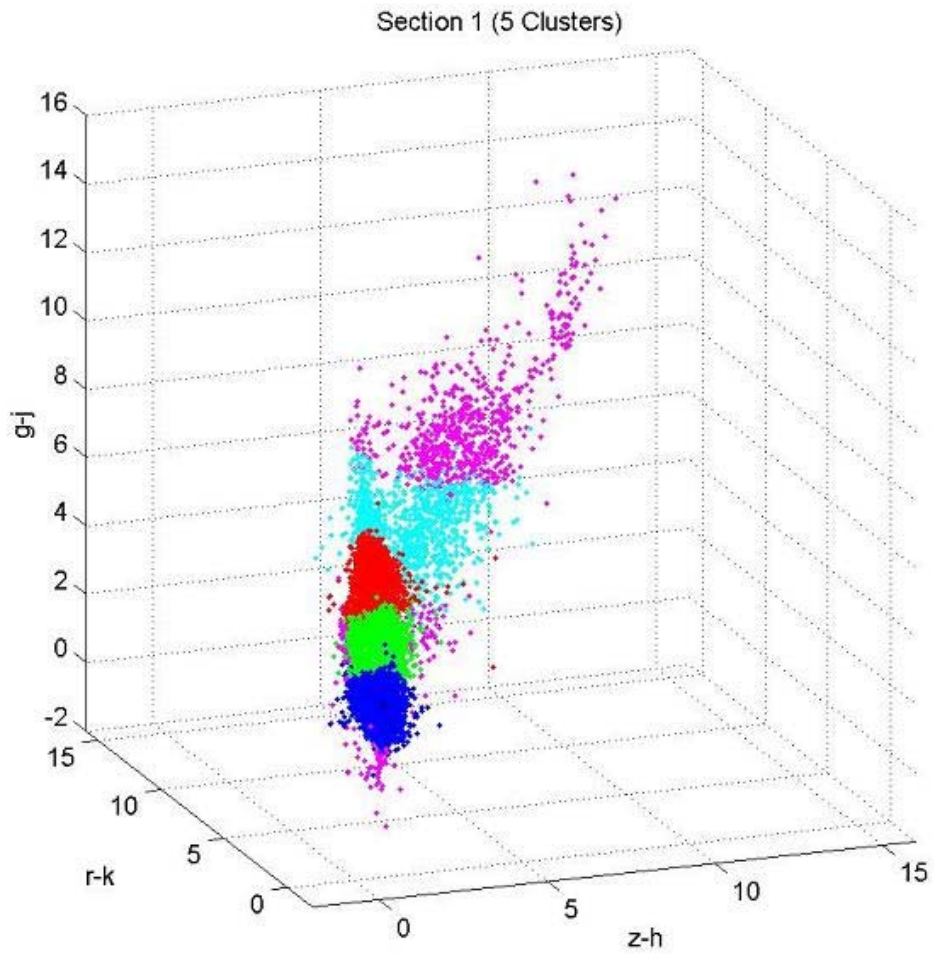
Petabytes!

We cannot even look through all our Terabyte data now.

What can we do when we have 1000's of Terabytes?



Automated hyperspace cluster analysis



Sample Machine Learning Applications for LSST:

***Automated Feature Extraction*:** Real-time identification of artifacts and transients in direct and difference images.

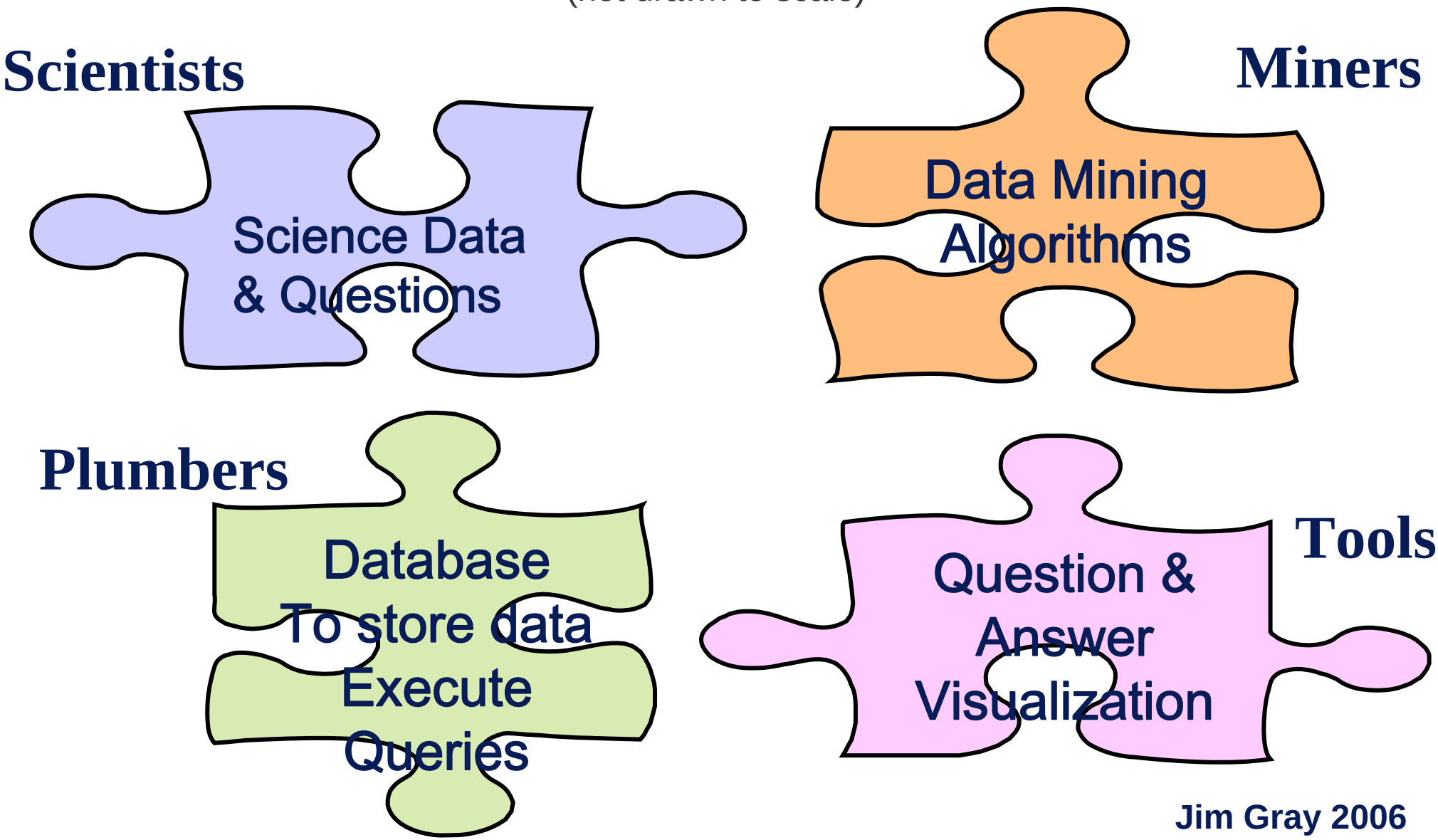
***Classifiers*:** Automated classification of celestial objects based on temporal and spectral properties.

***Anomaly Detection*:** Real-time recognition of important deviations from normal behavior for persistent sources.

http://www.thinkingtelescopes.lanl.gov/

What's needed?

(not drawn to scale)



Jim Gray 2006

Data Mining Research Challenge Areas:

- scalability (at petabytes scales) of existing machine learning and data mining algorithms
- development of grid-enabled parallel data mining algorithms
- multi-resolution methods for exploration of petascale databases
- visual data mining algorithms for visual exploration of the massive databases
- indexing of multi-attribute multi-dimensional astronomical databases
- rapid querying of petabyte databases

Scaleable Systems

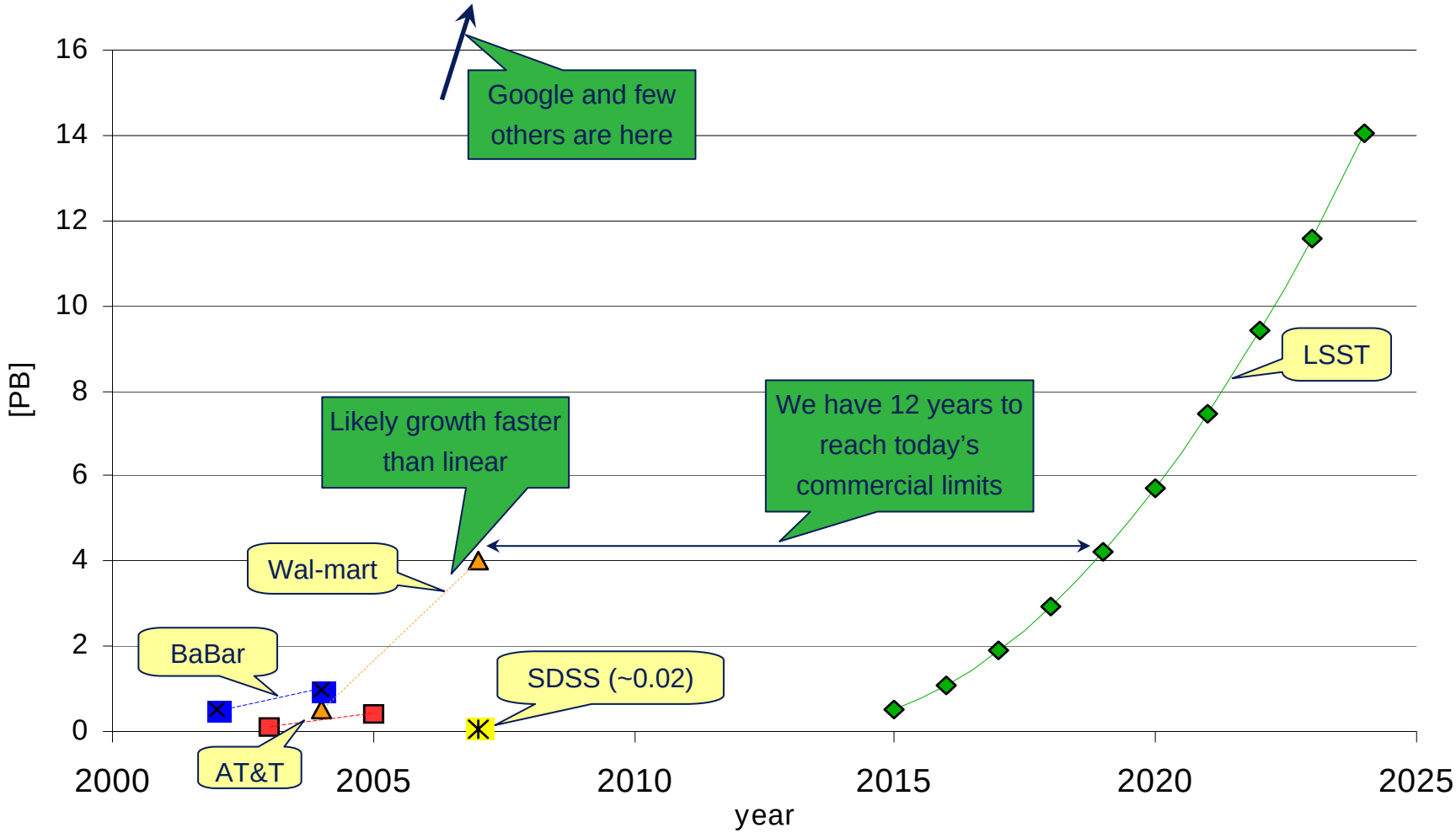
Scale UP

- Scale UP: **grow by adding components to a single system.**
- Scale Out**: grow by adding more systems.**

Scale OUT

Jim Gray 2006

Large Database Systems



* *All numbers based on publicly available data*

New paradigm

As the generation, analysis, communication, and preservation, of data are undergoing profound changes, scientific and engineering research is being similarly transformed.

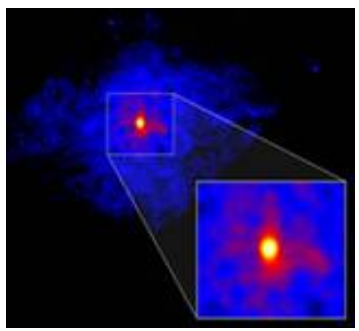
Data ➔ Information ➔ Knowledge ➔

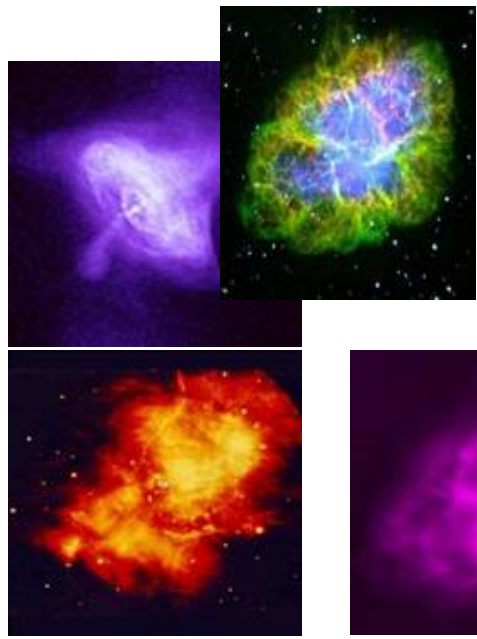
Understanding / Wisdom!

Szalay's Law:
The utility of N comparable datasets is N^2

- **Metcalf's law applies to telephones, fax, Internet.**
- **Szalay argues as follows:**
  Each new dataset gives new information
  2-way combinations give new information.
- **Example: Combine these 3 datasets**
  - **(ID, zip code)**
  - **(ID, birth day)**
  - **(ID, height)**
- **Other example:**
  quark star:

Chandra Xray +
Hubble optical,
+600 year old records..
Drake, J. J. et al.
Is RX J185635-375 a Quark Star?.
***Preprint*, (2002).**





Crab star
1053 AD

X-ray, optical, infrared, and radio views of the nearby Crab Nebula, which is now in a state of chaotic expansion after a supernova explosion first sighted in 1054 A.D. by Chinese Astronomers.

Jim Gray 2006

Publishing Data

Roles	*Traditional*	*Emerging*
Authors	**Scientists**	**Collaborations**
Publishers	**Journals**	**Project web site**
Curators	**Libraries**	**Data+Doc Archives**
Archives	**Archives**	**Digital Archives**
Consumers	**Scientists**	**Scientists**

Jim Gray 2006

New Issues

- How do we ensure data integrity?
- Who owns research data?
- How to treat the algorithms that were used?
- How should standards for access evolve?
- How do we assure quality metadata?
- Open access? [open data, open source]
- Peer review of petascale projects

Unimaginable opportunity

far from the end

a dichotomy

- **In the scientific method, high value is placed on negative findings.**
  (Theories can be disproved, not proved)

- **Current science culture places little value on “null” results.**
  (Journal articles finding no effect are valued less than articles reporting a positive effect)