the logic of exclusions

"If our motorboat engines were as erratic as our deliberate intellectual efforts, most of us would not get home for supper."

- Platt, *Science*, 1964

Michael J Keiser, PhD
Assistant Professor, UCSF

keiser@keiserlab.org
Bakar Computational Health Sciences Institute
Kavli Institute for Fundamental Neuroscience
Institute for Neurodegenerative Diseases

2019-09 NASEM Workshop

Enhancing Scientific Reproducibility through Transparent Reporting

16 October 1964, Volume 146, Number 3642

SCIENCE

Strong Inference

Certain systematic methods of scientific thinking may produce much more rapid progress than others.

John R. Platt

Scientists these days tend to keep up a polite fiction that all science is equal. Except for the work of the misguided opponent whose arguments we happen to be refuting at the time, we speak as though every scientist's field and methods of study are as good as every other scientist's, and perhaps a little better. This keeps us all cordial when it comes to recommending each other for government grants.

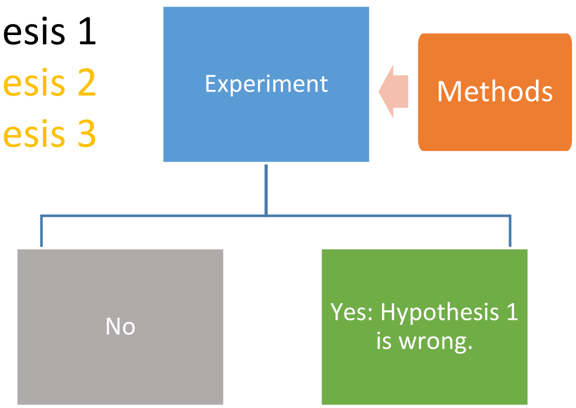
But I think anyone who looks at the matter closely will agree that some in scientific advance is an intellectual one. These rapidly moving fields are fields where a particular method of doing scientific research is systematically used and taught, an accumulative method of inductive inference that is so effective that I think it should be given the name of "strong inference." I believe it is important to examine this method, its use and history and rationale, and to see whether other groups and individuals might learn to adopt it profitably in their own scien-

"nature" or the experimental outcome chooses—to go to the right branch or the left; at the next fork, to go left or right; and so on. There are similar branch points in a "conditional computer program," where the next move depends on the result of the last calculation. And there is a "conditional inductive tree" or "logical tree" of this kind written out in detail in many first-year chemistry books, in the table of steps for qualitative analysis of an unknown sample, where the student is led through a real problem of consecutive inference: Add reagent A; if you get a red precipitate, it is subgroup alpha and you filter and add reagent B; if not, you add the other reagent, B′; and so on.

On any new problem, of course, inductive inference is not as simple and certain as deduction, because it involves reaching out into the unknown. Steps 1 and 2 require intellectual inventions, which must be cleverly chosen so that hypothesis, experiment, outcome, and exclusion will be related in a rigorous syllogism; and the question of how to generate such

strong inference follows a systematic & transparent recipe

Hypothesis 1
Hypothesis 2
Hypothesis 3



- 1) Devising alternative hypotheses
- 2) Devising crucial experiment(s),
  - with alternative possible outcomes, each of which will, as nearly as possible,
  - exclude one or more of the hypotheses
- 3) Carrying out the experiment so as to get a clean result

- 1') Recycling the procedure, making
  - sub-hypotheses or
  - sequential hypotheses to refine the possibilities that remain; and so on.

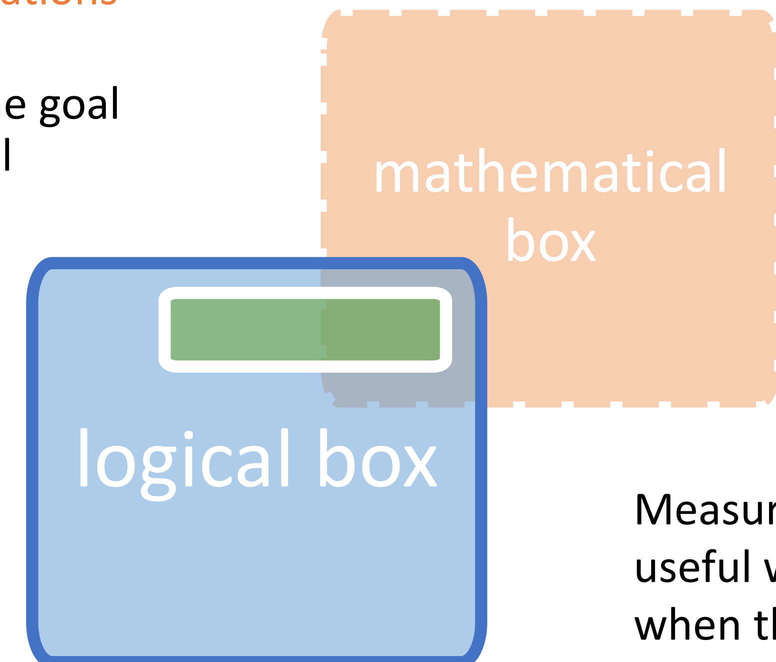
Platt, *Science*, 1964

which measurements are effective?

- Errors of reasoning:
    - We substitute correlations for causal studies.
    - Numbers become the goal instead of the crucial experiment.

elegant
fine-grained
but flimsy

coarse
qualitative
strong



Measurements are useful when & only when they are related to disproof.

"Many—perhaps most—of the great issues of science are qualitative, not quantitative."

Platt, *Science*, 1964

A model which cannot be mortally endangered cannot be alive.

- W. A. H. Rushton meets machine learning

Adversarial Controls for Scientific Machine Learning

ACS Chem Biol. 2018 Oct 19. Chuang KV, Keiser MJ.

1. **Opening the Black Box**
   Does the model make scientific sense?

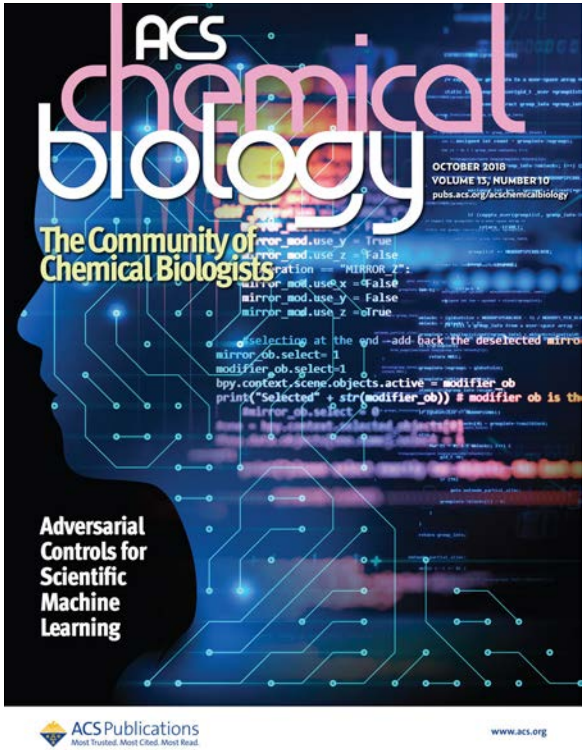
   Platt's logical box 1964

2. **The Method of Multiple Models**
   Is a confounding variable driving the prediction?

   Chamberlin 1897, Bacon 1620

3. **Outperforming the Straw Model**
   Does it break when you remove what matters?

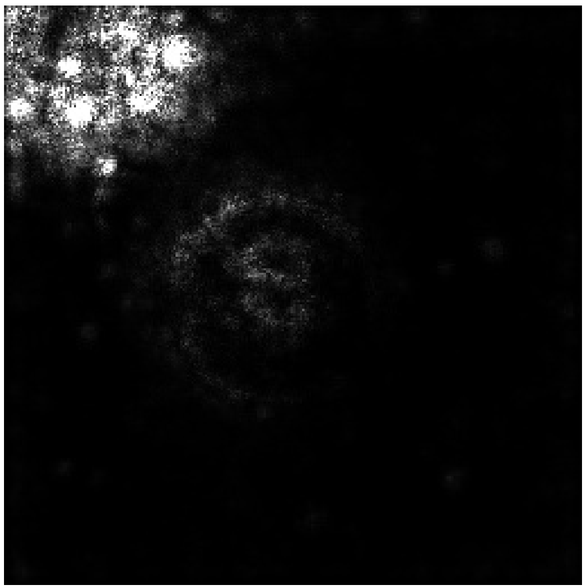
   Langley's straw man 1988
   Popper 1963



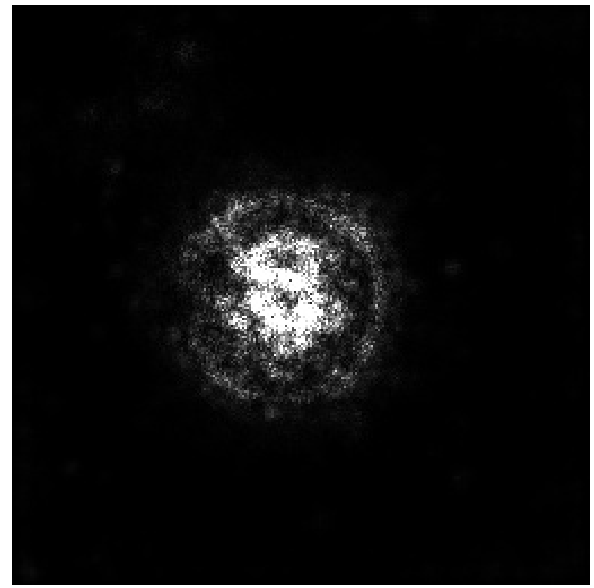
- Langley, P. (1988) Machine Learning as an Experimental Science. Mach. Learn. 3 (1), 5−8.
- Science as Falsification (Popper, *Conjectures and Refutations*, 1963)
- *The method of multiple hypotheses* (Chamberlin, 1897)

Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline

Nat Commun. 2019 May 15. Tang Z, et al.



Diffuse



Cored

https://doi.org/10.1101/454793
https://github.com/keiserlab/plaquebox-paper
https://doi.org/10.5281/zenodo.1470797

keiser lab @ UCSF

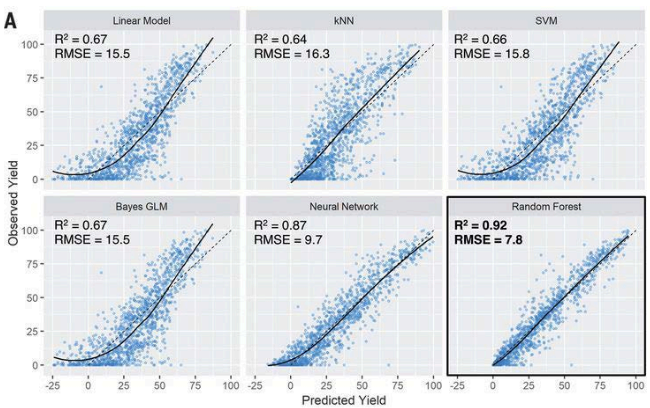
1 - black box

Comment on "Predicting reaction performance in C-N cross-coupling using machine learning"

Science. 2018 Nov 16. Chuang KV, Keiser MJ.

Science. 13 Apr 2018. Ahneman et al.

Prediction of chemical reaction yields



SHARE

23

2

IN THE PIPELINE

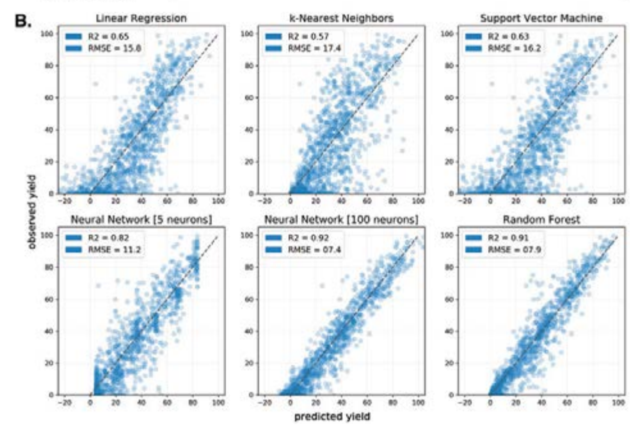
Derek Lowe's commentary on drug discovery and the pharma industry. An editorially independent blog from the publishers of *Science Translational Medicine*. All content is Derek's own, and he does not in any way speak for his employer.

By Derek Lowe

CHEMICAL NEWS

Machine Learning: Be Careful What You Ask For

By **Derek Lowe** | 20 November, 2018



Using **random barcodes** instead of chemical features

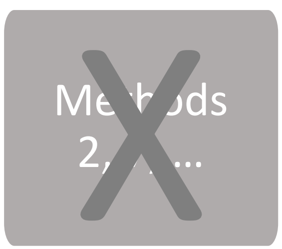
https://github.com/keiserlab/comments

keiser lab @ UCSF

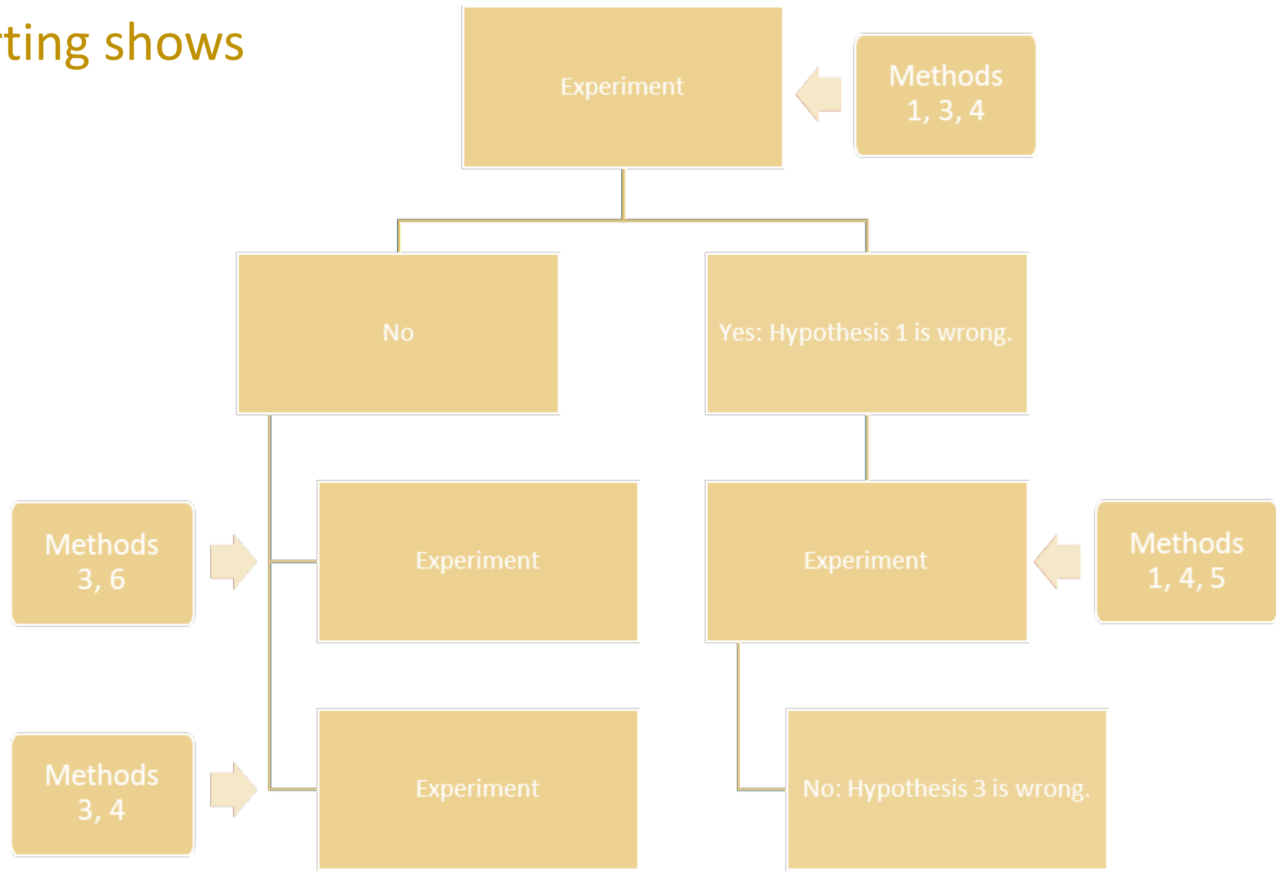
3 - straw model

how should we choose scientific reporting standards?

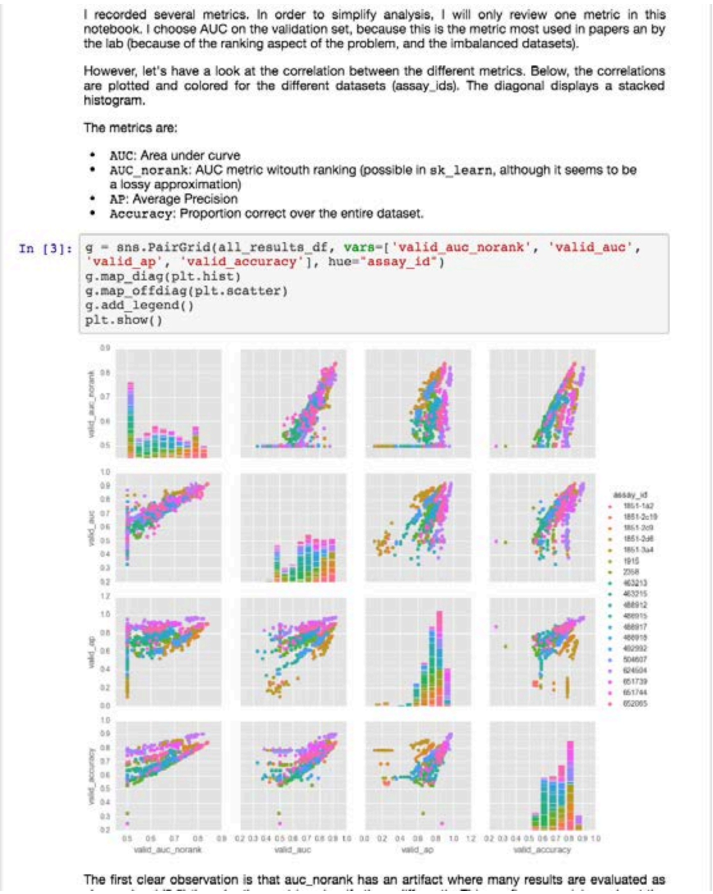
- Transparent reporting shows a chain of precise induction of how nature works.



- This is the sole yardstick of their effectiveness.



we can share the chain itself



I recorded several metrics. In order to simplify analysis, I will only review one metric in this notebook. I choose AUC on the validation set, because this is the metric most used in papers an by the lab (because of the ranking aspect of the problem, and the imbalanced datasets).

However, let's have a look at the correlation between the different metrics. Below, the correlations are plotted and colored for the different datasets (assay_ids). The diagonal displays a stacked histogram.

The metrics are:

- `AUC`: Area under curve
- `AUC_norank`: AUC metric witouth ranking (possible in `sk_learn`, although it seems to be a lossy approximation)
- `AP`: Average Precision
- `Accuracy`: Proportion correct over the entire dataset.

In [3]:
```
g = sns.PairGrid(all_results_df, vars=['valid_auc_norank', 'valid_auc',
'valid_ap', 'valid_accuracy'], hue="assay_id")
g.map_diag(plt.hist)
g.map_offdiag(plt.scatter)
g.add_legend()
plt.show()
```


The first clear observation is that auc_norank has an artifact where many results are evaluated as

- Registered reports
  - https://osf.io/rr/
- Version control (git) & data (zenodo)
  - https://github.com & https://zenodo.org
- Show logic in Jupyter notebook/lab
  - http://jupyter.org
- Save the environment (conda)
  - https://github.com/conda/conda
- Use makefiles &/or workflow tools
  - http://kbroman.org/minimal_make
  - https://github.com/pditommaso/awesome-pipeline
- *10 Rules for Reproducible Research*
  - PLoS Comp Biol 2013
  - http://bit.ly/2bhhSQx

one proposal-

scientific (ai) red team

& regular in-lab **code review**


Albert Young
Research Fellow (MS4) // medical program
— *Albert is a 4th year UCSF medical student interested ...*


Ben Wong
Systems and Infrastructure Admin
— *Ben keeps the trains running and the GPU fans spinnin...*


Daniel Wong
Grad Student // bioinformatics
— *Daniel studied Computer Science and Biochemistry at U...*


Edward Elhauge, MPH
Specialist


Elena Caceres
Grad Student; NSF Fellow; HHMI Gilliam Fellow // bioinformatics
— *Elena graduated from UCSD with a B.Sc. in molecular b...*


Garrett Gaskins
Grad Student; Genentech Fellow; Hillblom Fellow // bioinformatics
— *High-content screening across varied cells, condition...*


Jacob Pfau, M1
Research Data Analyst; QBI Bold & Basic Fellow
— *My interests span the theory and applications of mach...*


Jessica McKinley, PhD
Postdoctoral Scholar
— *Jessica graduated from UC Riverside with a PhD in Com...*


Kangway Chuang, PhD
Arnold O. Beckman Postdoctoral Fellow


Laura Gunsalus
Grad Student // bioinformatics
— *I'm a iPQB bioinformatics graduate student interested...*


Leo Gendelev
Grad Student; Fletcher Jones Fellow // biophysics
— *Leo came into the micro-world of molecules and cells ...*


Luca Ponzoni, PhD
Postdoctoral Scholar
— *Luca obtained his PhD in Physics and Chemistry of Bio...*


Michael Keiser, PhD
Assistant Professor; CZI Ben Barres Investigator; Allen Distinguished Investigator // pharm chem; bts; bchsc kifn; ind
— *Michael is a Chan Zuckerberg Initiative Ben Barres In...*


Nick Mew, MS
Specialist
— *Coming from a software engineering and computer scien...*


Will Connell
Grad Student // pspg
— *Currently, clinicians practice medicine on a populati...*


Wren Saylor
Grad Student // bioinformatics
— *Wren studied general biology at Hampshire College. Sh...*

CHAN ZUCKERBERG INITIATIVE

THE PAUL G. ALLEN FRONTIERS GROUP

NIH NIGMS

NSF

keiserlab.org

alumni


Alexandre Fassio
Graduate Student // Federal University of Minas Gerais
— *Visiting CAPES Scholar, 2018-2019. Alexandre was a vi...*


Amanda Li, PhD
Decentralized Consensus Fellow // Insight Data Science
— *Postdoctoral Fellow at UCSF and the Accelerating Ther...*


Cristina Melero, MS
— *Researcher and Lab Manager, 2014-2016.*


Jizhou Yang, MS
Bioinformatics Engineer // Genentech
— *Data science intern, 2017-2018.*


Michael Mysinger, PhD
Principal Scientist // Atomwise
— *2015-2016.*


Teague Sterling, MS
Principal Engineer, Research Informatics // BioMarin Pharmaceutical


Ziqi Tang
Master's Student, Computer Science // Georgia Tech

UCSF

University of California
San Francisco