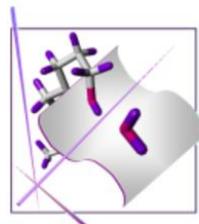
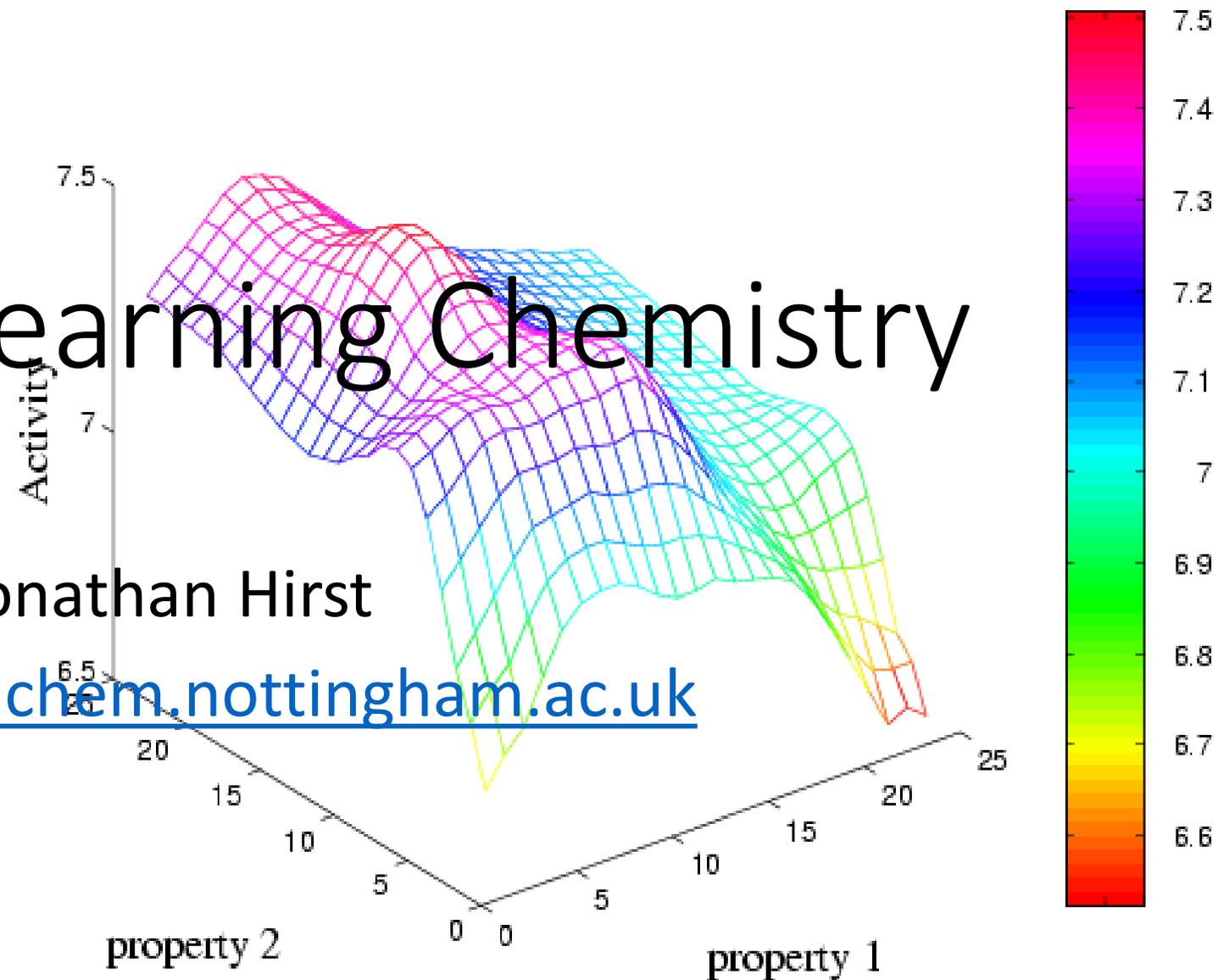




Machines Learning Chemistry

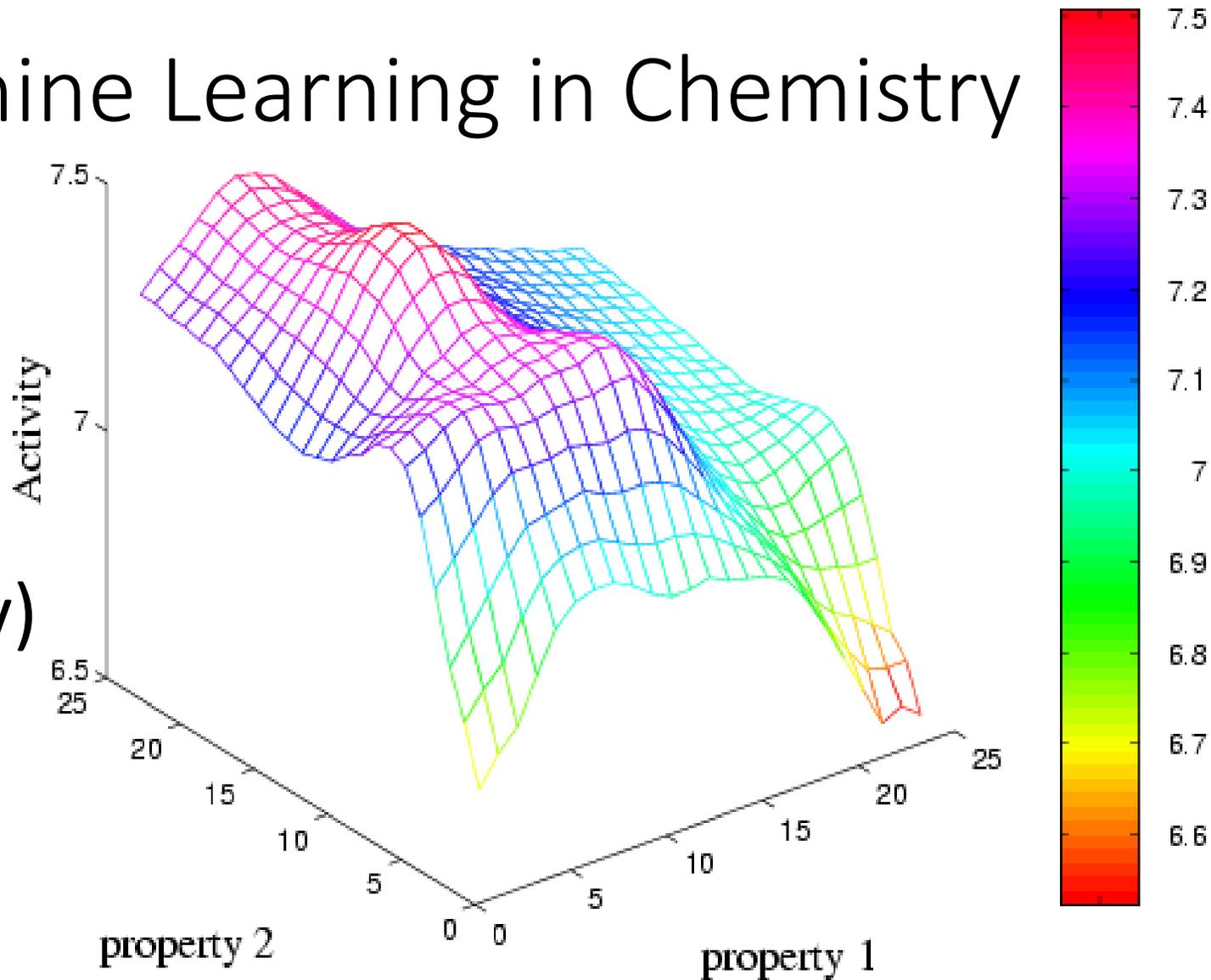
Jonathan Hirst

<http://comp.chem.nottingham.ac.uk>

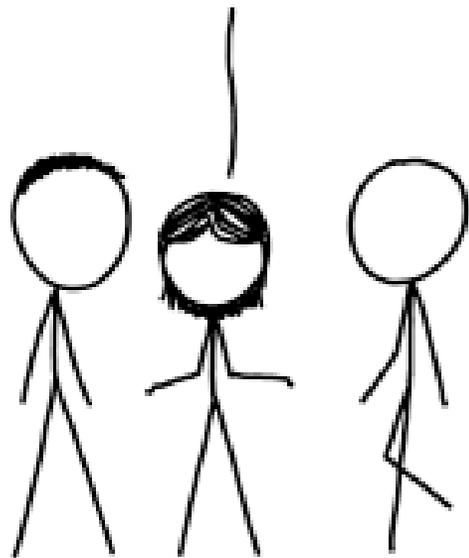


Overview of Machine Learning in Chemistry

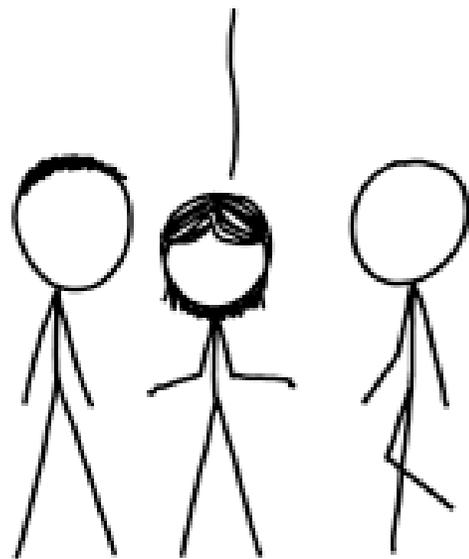
- Excitement (general)
- Historical context
- Methods
- Excitement (chemistry)
- Our recent work
 - Steve Oatley
- Conclusions



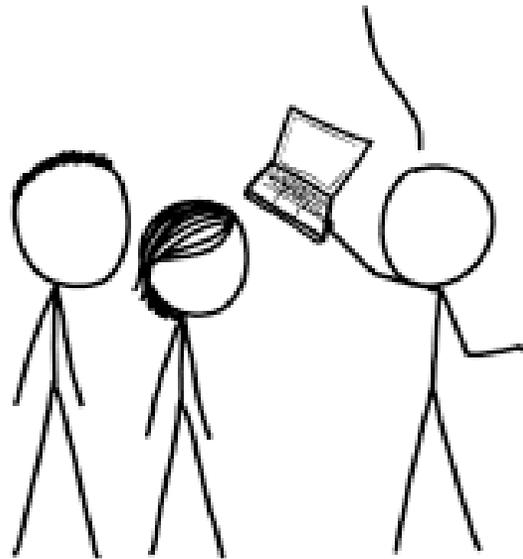
OUR FIELD HAS BEEN
STRUGGLING WITH THIS
PROBLEM FOR YEARS.



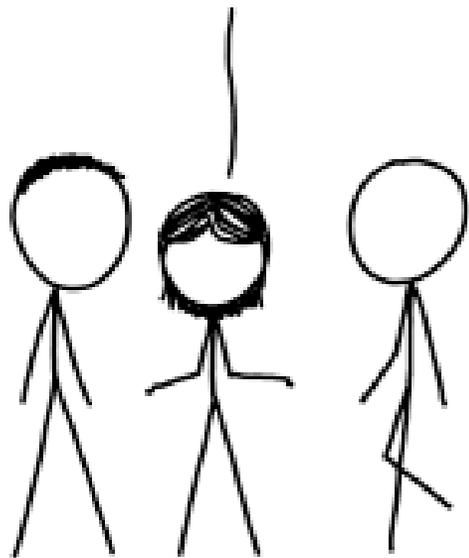
OUR FIELD HAS BEEN
STRUGGLING WITH THIS
PROBLEM FOR YEARS.



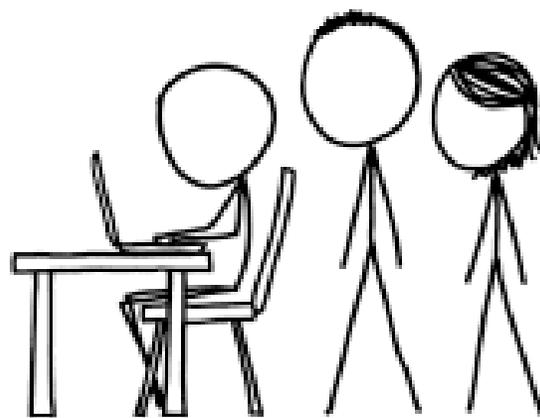
STRUGGLE NO MORE!
I'M HERE TO SOLVE
IT WITH *ALGORITHMS!*



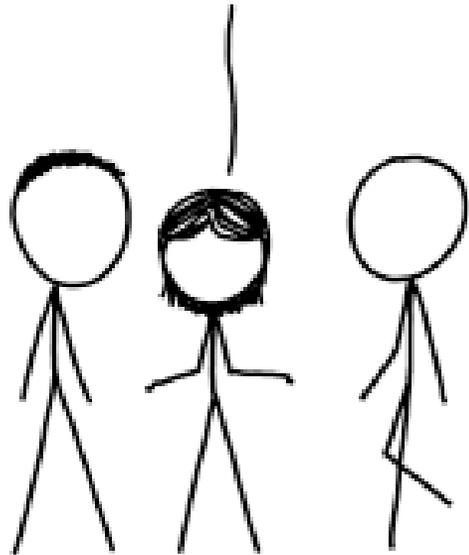
OUR FIELD HAS BEEN
STRUGGLING WITH THIS
PROBLEM FOR YEARS.



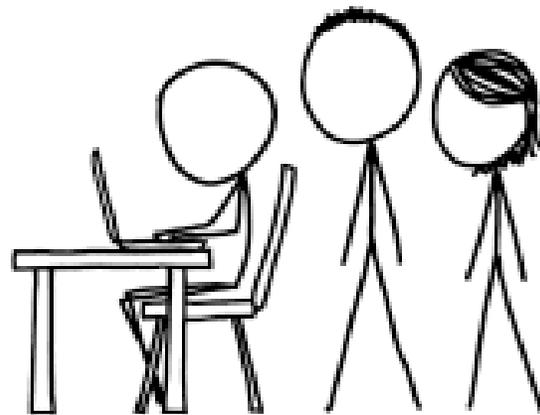
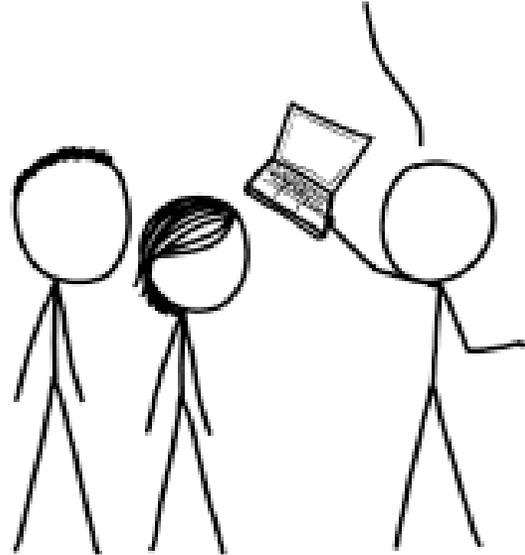
STRUGGLE NO MORE!
I'M HERE TO SOLVE
IT WITH *ALGORITHMS!*



OUR FIELD HAS BEEN STRUGGLING WITH THIS PROBLEM FOR YEARS.



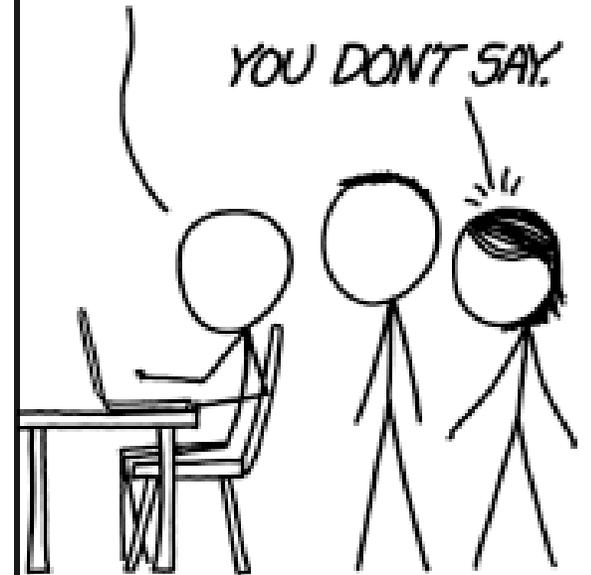
STRUGGLE NO MORE! I'M HERE TO SOLVE IT WITH *ALGORITHMS!*



SIX MONTHS LATER:

WOW, THIS PROBLEM IS REALLY HARD.

YOU DON'T SAY!



Excitement

- “Here I am just sitting in this house and I’m able to predict a cure to measles,” co-founder of Atomwise Alex Levy tells me [Sarah Buhr] <https://techcrunch.com/2015/03/06/y-combinator-backed-atomwise-discovers-drugs-for-diseases-that-dont-even-exist-yet/>
- “If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future.” Ng, A. (2016). What Artificial Intelligence Can and Can’t Do Right Now. *Harvard Business Review*.

ImageNet Classification with Deep Convolutional Neural Networks

By: Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E.

COMMUNICATIONS OF THE ACM Volume: 60 Issue: 6 Pages: 84-90 Published: JUN 2017



Free Full Text from Publisher

View Abstract ▼

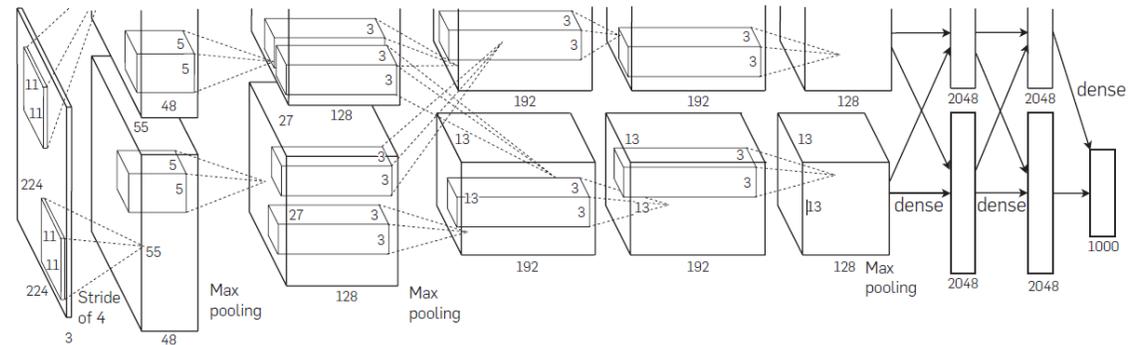
Times Cited: 16,653
(from Web of Science Core Collection)



Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0%, respectively, which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully connected layers we employed a recently developed regularization method called “dropout” that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

Figure 2. An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 290,400–186,624–64,896–64,896–43,264–4096–4096–1000.



The original version of this paper was published in the *Proceedings of the 25th International Conference on Neural Information Processing Systems* (Lake Tahoe, NV, Dec. 2012), 1097–1105.

LEARNING REPRESENTATIONS BY BACK-PROPAGATING ERRORS

By: RUMELHART, DE (RUMELHART, DE); HINTON, GE (HINTON, GE); WILLIAMS, RJ (WILLIAMS, RJ)

NATURE

Volume: 323 Issue: 6088 Pages: 533-536

DOI: 10.1038/323533a0

Published: OCT 9 1986

Citation Network

In Web of Science Core Collection

7,212

Times Cited

SUPPORT-VECTOR NETWORKS

By: CORTES, C (CORTES, C); VAPNIK, V (VAPNIK, V)

MACHINE LEARNING

Volume: 20 Issue: 3 Pages: 273-297

DOI: 10.1023/A:1022627411411

Published: SEP 1995

Citation Network

In Web of Science Core Collection

16,864

Times Cited

Deep learning

By: LeCun, Y (LeCun, Yann)^[1,2]; Bengio, Y (Bengio, Yoshua)^[3]; Hinton, G (Hinton, Geoffrey)^[4,5]

[View Web of Science ResearcherID and ORCID](#)

NATURE

Volume: 521 Issue: 7553 Pages: 436-444

DOI: 10.1038/nature14539

Published: MAY 28 2015

Citation Network

In Web of Science Core Collection

8,673

 Highly Cited Paper

Times Cited

 Create Citation Alert



AlphaGo

nature



2017

Go – a typical game has a possible 10^{800} moves
Monte Carlo Tree Search and deep neural networks

2019

SHARE

PERSPECTIVE | COMPUTER SCIENCE



AI surpasses humans at six-player poker



Alan Blair, Abdallah Saffidine

+ See all authors and affiliations



Science 30 Aug 2019:
Vol. 365, Issue 6456, pp. 864-865
DOI: 10.1126/science.aay7774

Summary

Superhuman performance by artificial intelligence (AI) has been demonstrated in two-player, deterministic, zero-sum, perfect-information games (1) such as chess, checkers (2), Hex, and Go (3). Research using AI has broadened to include games with challenging attributes such as randomness, multiple players, or imperfect information. Randomness is a feature of dice games, and card games include the additional complexity that each player sees some cards that are hidden from others. These aspects more closely resemble real-world situations, and this research may thus lead to algorithms with wider applicability. On page 885 of this issue, Brown and Sandholm (4) show that a new computer player called Pluribus exceeds human performance for six-player Texas hold'em poker.

Historical context

ESCOM

PD3 019

1993

New approaches to QSAR: Neural networks and machine learning

Ross D. King, Jonathan D. Hirst and Michael J.E. Sternberg*

*Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, P.O. Box 123,
London WC2A 3PX, U.K.*

PD3 019



1993

AR: Neural networks and machine learning

D. Hirst and Michael J.E. Sternberg*

*al Cancer Research Fund, 44 Lincoln's Inn Fields, P.O. Box 123,
ndon WC2A 3PX, U.K.*

“My CPU is a neural-net processor; a learning computer” (1991)



Designing drugs by computer at Merck



1981

Designing drugs by computer at Merck

Hiller SA, Golender VE, Rosenblit AB.
Cybernetic methods of drug design. I.
Statement of the problem—The perceptron
approach. *Computers and Biomedical
Research*, 6, 411-421.

1973

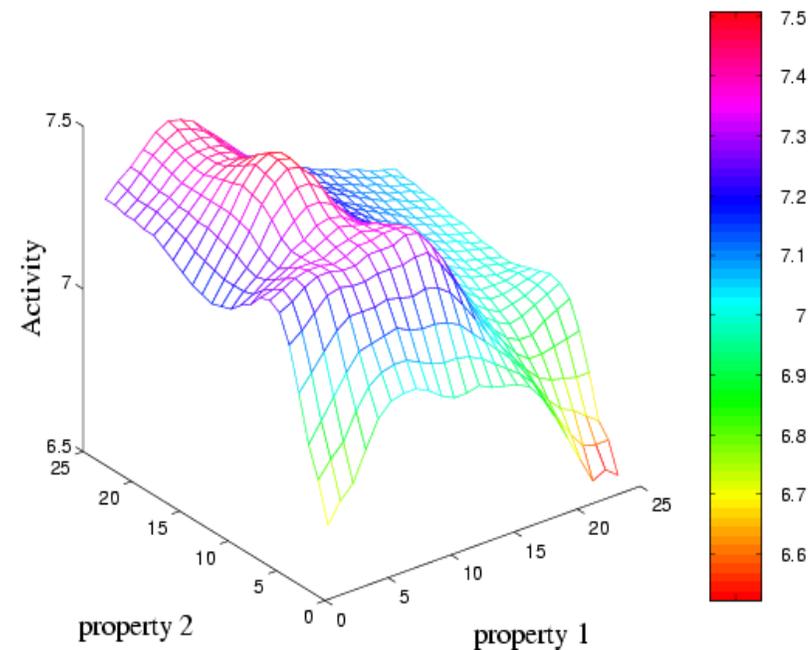
It is revealed that the problem of drug design which is at present coped with on a semi-intuitive basis may be interpreted in terms of modern pattern recognition theory as a problem of discriminating two classes of objects: the active and the inactive chemical compounds.

In the meantime two questions are essentially important: (1) the presentation of information on the structure of a chemical compound, i.e., the elaboration of terms for adequately describing the structure and (2) the selection of a recognition algorithm.

Methods

What is machine learning?

- ~~Magic~~
- Non-parametric regression
- Neural networks
- Support vector machines
- Random forests
- Inductive logic programming



Bruce CL *et al.* Contemporary QSAR Classifiers Compared. *J Chem Inf Model* **2007** 47, 219-227

Melville JL *et al.* Machine Learning in Virtual Screening. *Combin Chem High Thr Scr* **2009**, 12, 332-343

Non-parametric regressors

- Non-parametric estimates of a function appear as a linear combination of the measured responses, y ,

$$\hat{m}(x) = \sum_i w_i(x) y_i$$

- For example, the weights might be defined as Euclidean distances

$$w_i(x) = \frac{d_i(x)^{-1}}{\sum_j d_j(x)^{-1}}$$

Hirst, *J Med Chem* **39** 3526 (1996); Constans & Hirst, *J Chem Inf Comput Sci* **40** 452 (2000); McNeany & Hirst, *J Chem Inf Model* **45** 768 (2005)

Nadaraya-Watson kernel estimator

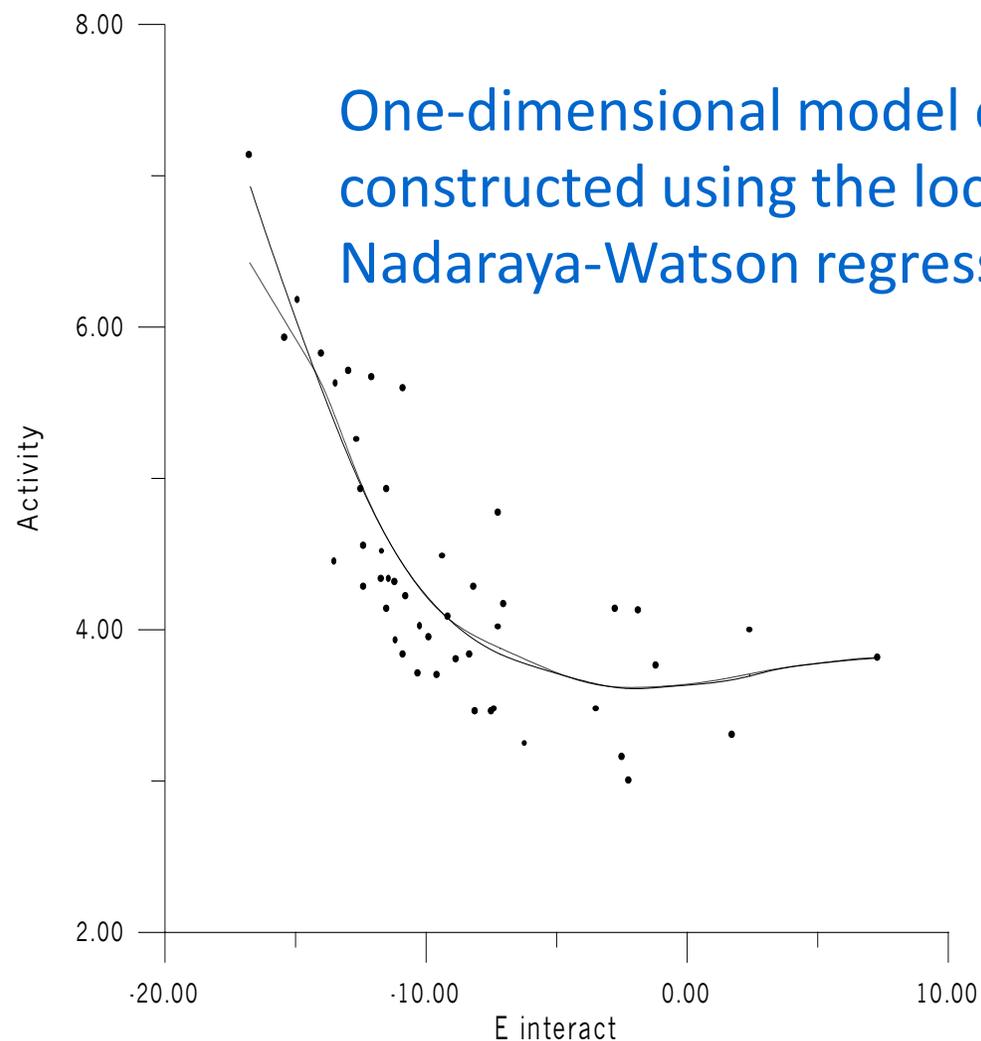
$$\hat{m}_{NW}(x) = \frac{\sum_{i=1}^N K_x\left(\frac{x-x_i}{h_x}\right) y_i}{\sum_{i=1}^N K_x\left(\frac{x-x_i}{h_x}\right)}$$

K is the **smoothing kernel**; h is the **bandwidth**

$$K(u/h) = (2\pi)^{-1/2} e^{-u^2/2h^2}$$

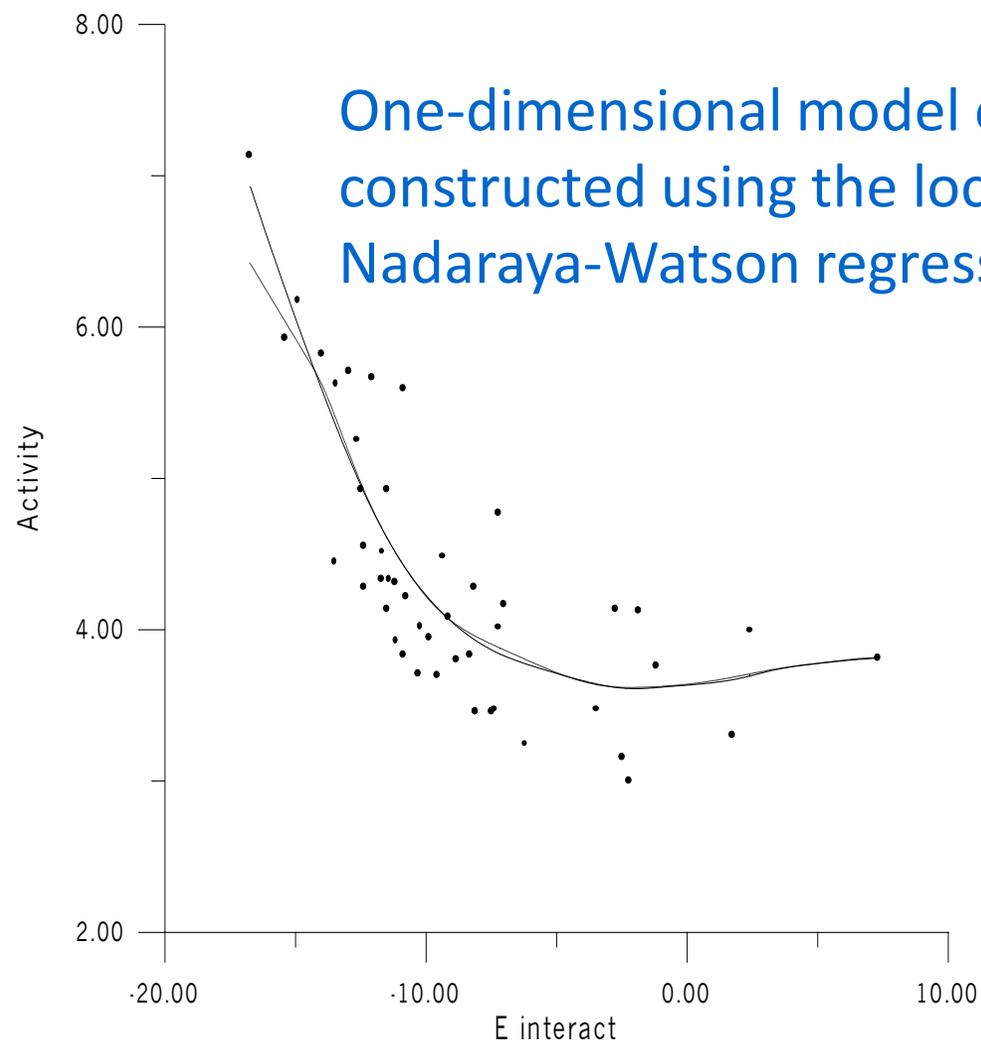
h determines the extent of smoothing

Hirst, *J Med Chem* **39** 3526 (1996); Constans & Hirst, *J Chem Inf Comput Sci* **40** 452 (2000);
McNeany & Hirst, *J Chem Inf Model* **45** 768 (2005)

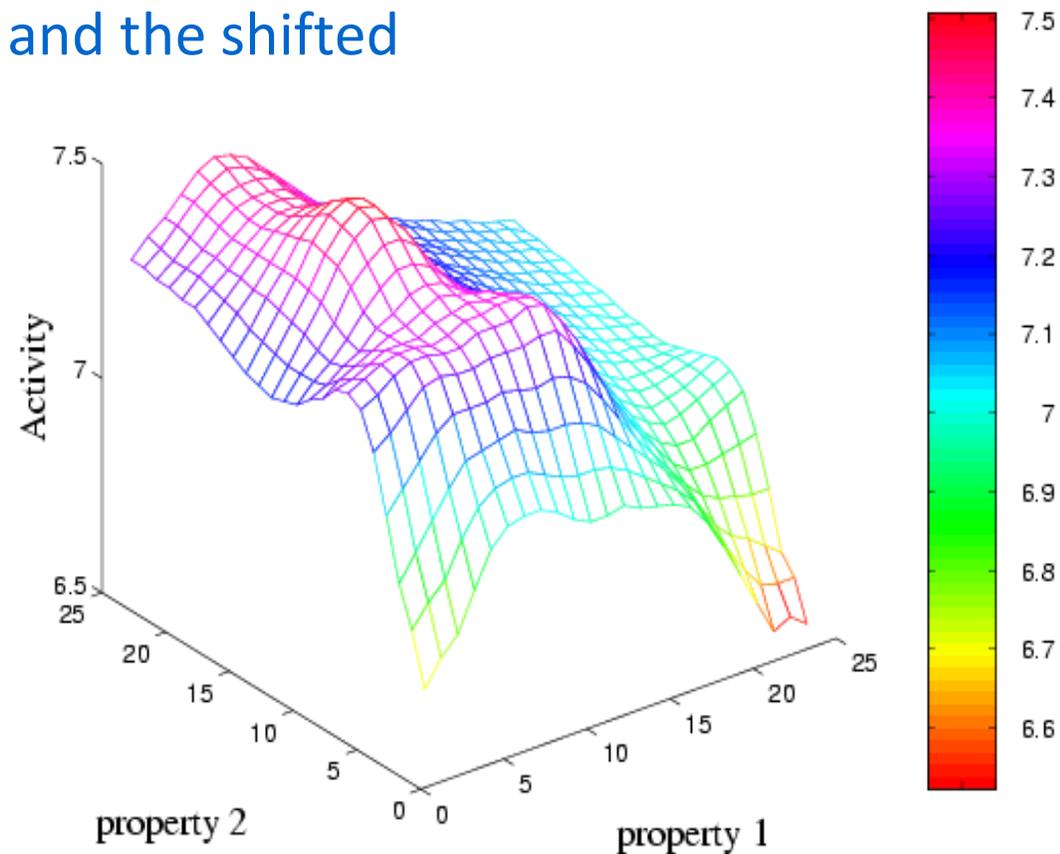


One-dimensional model of activity as a function of E_{interact} constructed using the local linear and the shifted Nadaraya-Watson regressors

Hirst, *J Med Chem* **39** 3526 (1996); Constans & Hirst, *J Chem Inf Comput Sci* **40** 452 (2000); McNeany & Hirst, *J Chem Inf Model* **45** 768 (2005)



One-dimensional model of activity as a function of E_{interact} constructed using the local linear and the shifted Nadaraya-Watson regressors



Hirst, *J Med Chem* **39** 3526 (1996); Constans & Hirst, *J Chem Inf Comput Sci* **40** 452 (2000); McNeany & Hirst, *J Chem Inf Model* **45** 768 (2005)

Excitement (chemistry)



COMPUTATIONAL CHEMISTRY

Is machine learning overhyped?

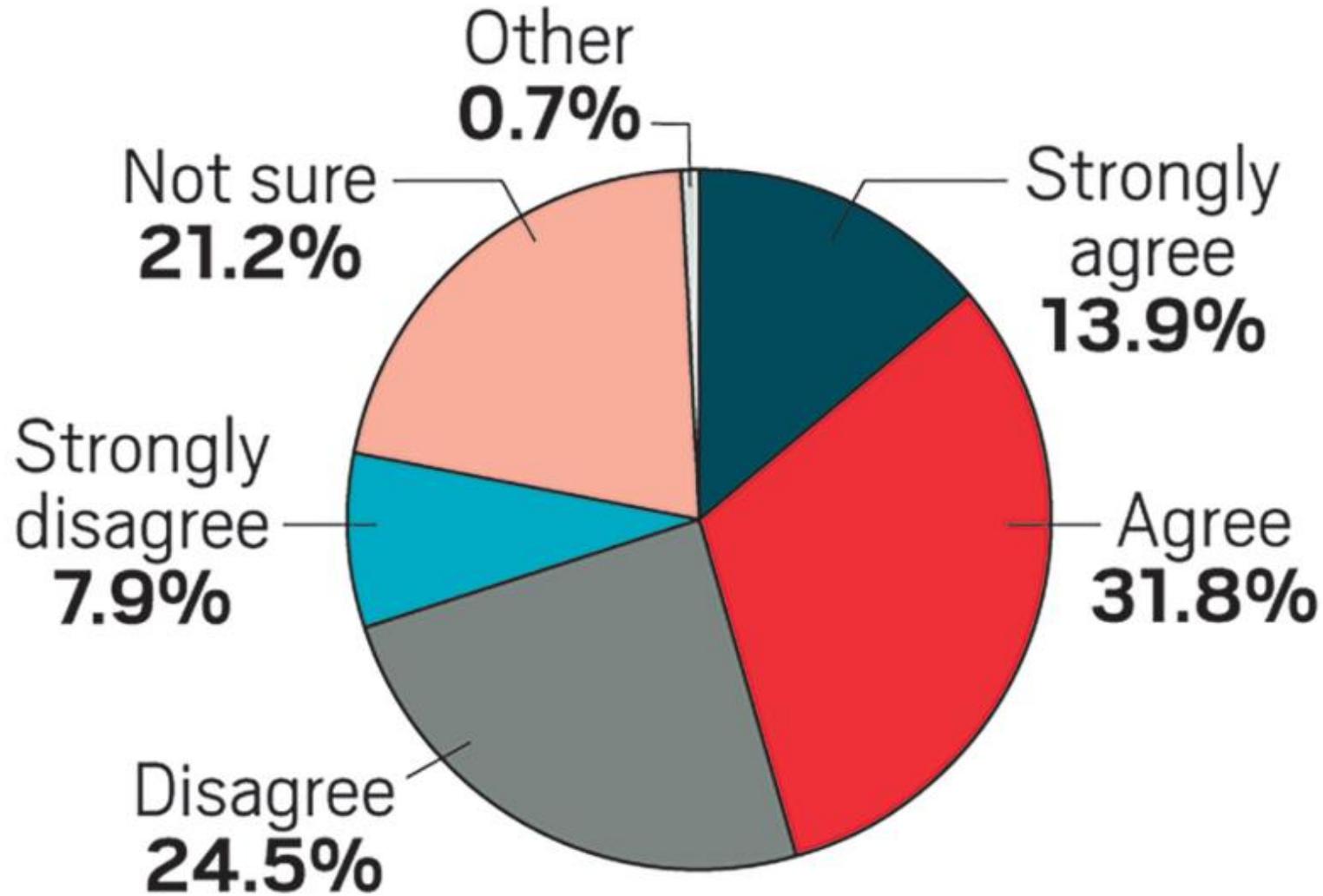
Chemists weigh in on the technique's possibilities and its pitfalls

by **Sam Lemonick**

AUGUST 27, 2018 | APPEARED IN **VOLUME 96, ISSUE 34**

<https://cen.acs.org/physical-chemistry/computational-chemistry/machine-learning-overhyped/96/i34>

Do you think machine learning is overhyped?



Total respondents = 151

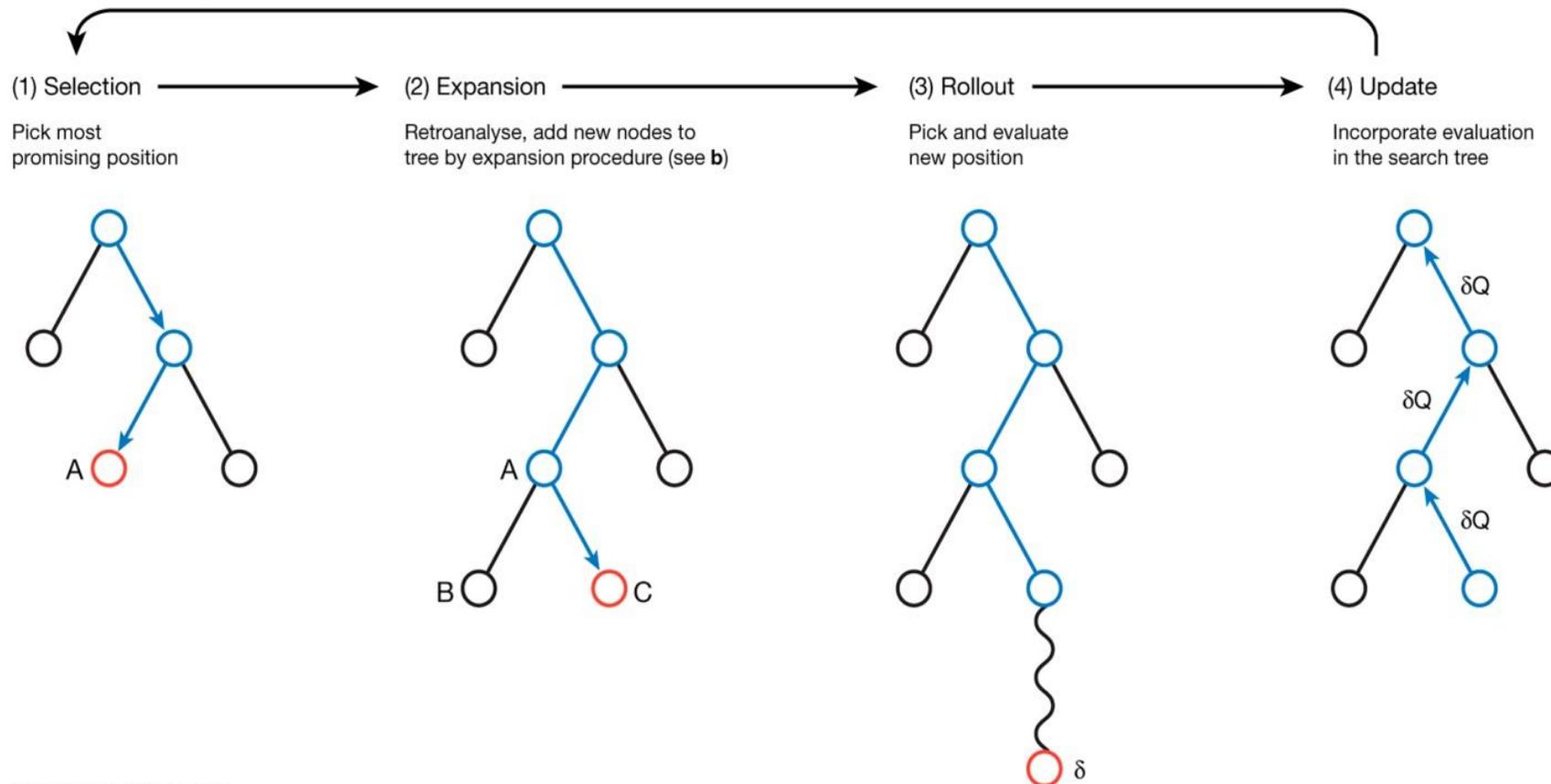
The combination of the **Monte Carlo Tree Search and deep neural networks** is extremely well suited to planning chemical syntheses - so-called retrosyntheses - with unprecedented efficiency.

Segler: "Retrosynthesis is the ultimate discipline in organic chemistry. Chemists need years to master it - just like with chess or Go.

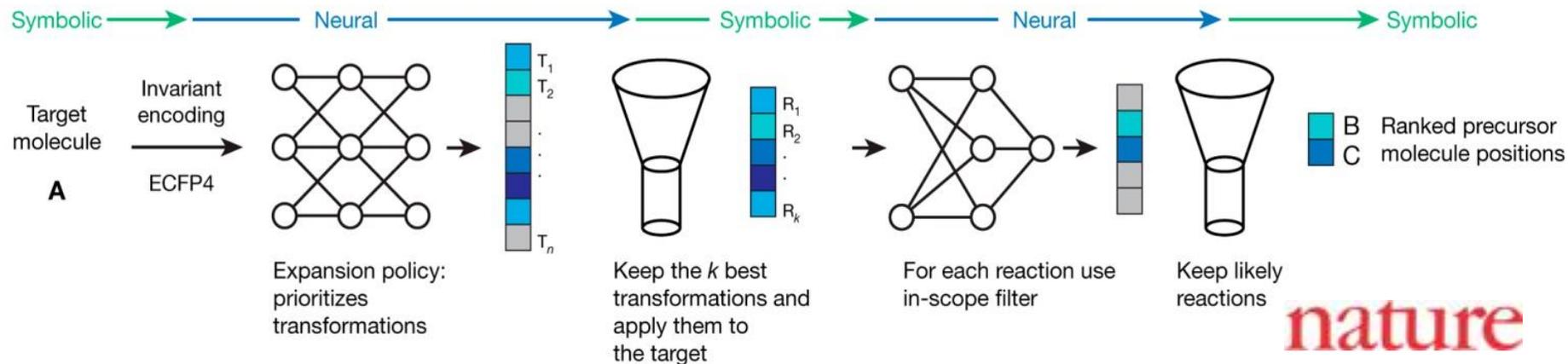
In addition to straightforward expertise, you also need a goodly portion of intuition and creativity for it. So far, everyone assumed that computers couldn't keep up without experts programming in tens of thousands of rules by hand.

What we have shown is that the machine can, by itself, learn the rules and their applications from the literature available."

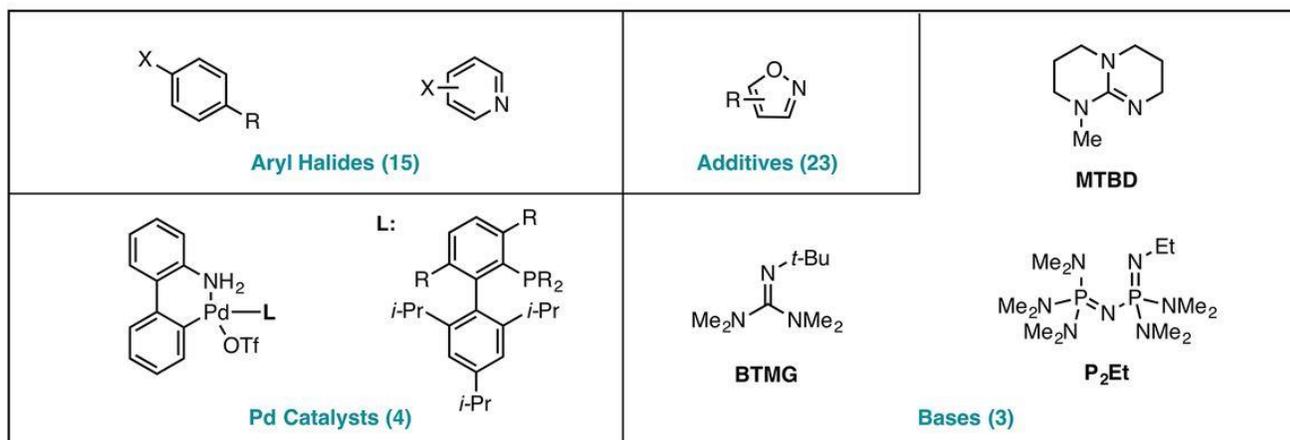
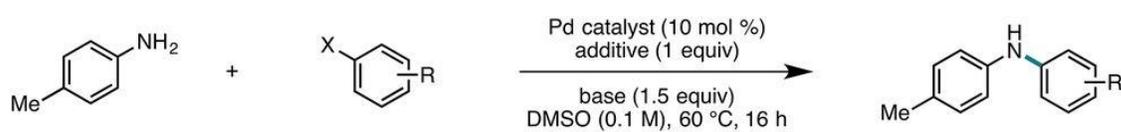
https://www.eurekalert.org/pub_releases/2018-03/uom-csw032918.php



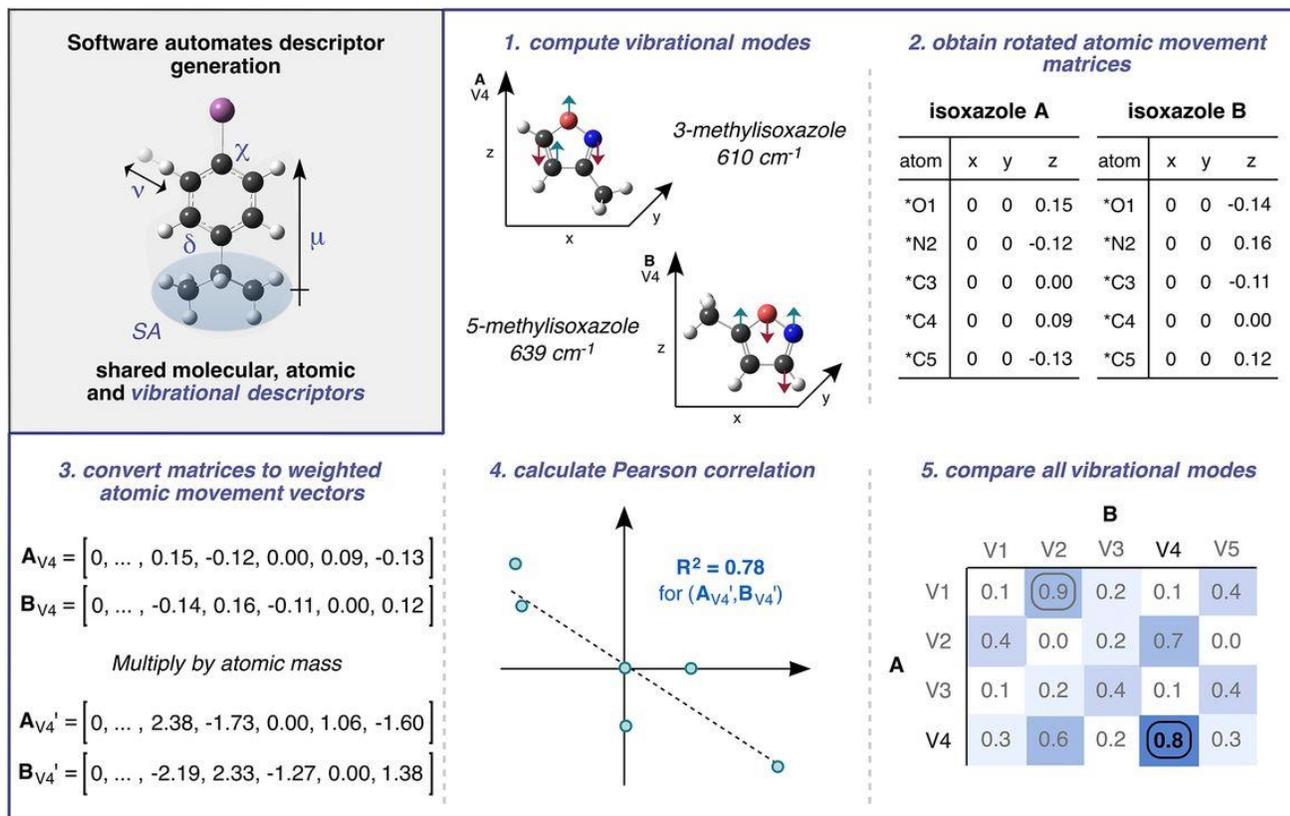
b Expansion procedure



A



B



Application of machine learning to reaction prediction

Ahneman et al.
Science 2018;
 science.aar5169



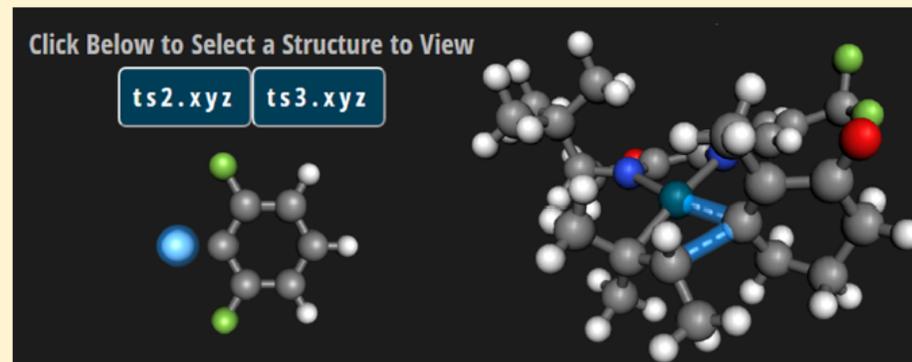
AARON: An Automated Reaction Optimizer for New Catalysts

Yanfei Guan,[†] Victoria M. Ingman,[§] Benjamin J. Rooks,[†] and Steven E. Wheeler^{*,†,§} 

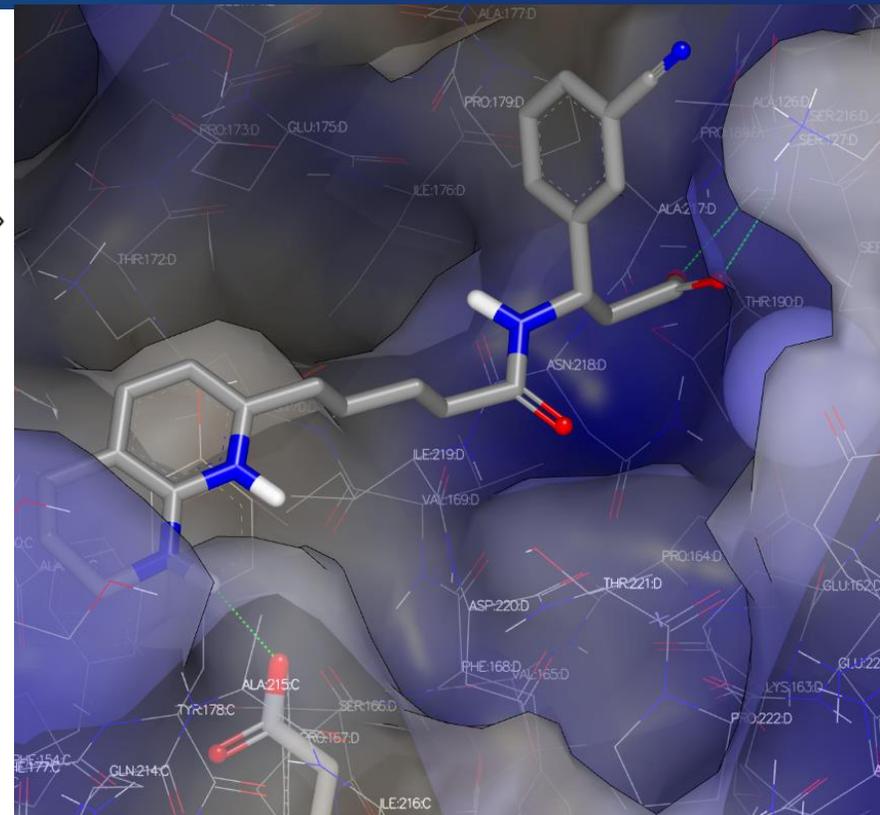
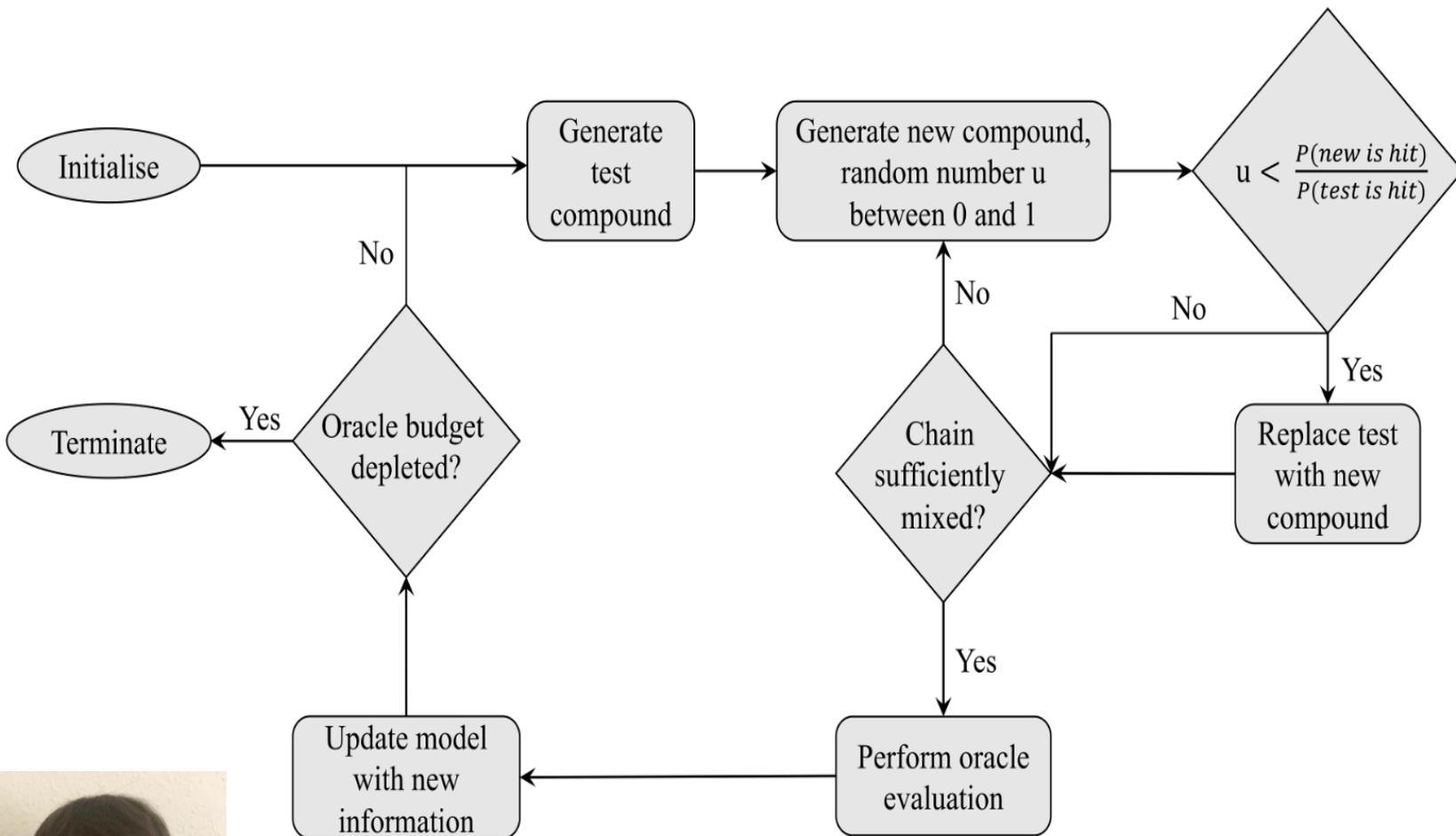
[†]Department of Chemistry, Texas A&M University, College Station, Texas 77842, United States

[§]Center for Computational Quantum Chemistry, Department of Chemistry, University of Georgia, Athens, Georgia 30602, States

ABSTRACT: We describe an open-source computational toolkit (AARON: An Automated Reaction Optimizer for New catalysts) that automates the quantum mechanical geometry optimization and characterization of the transition state and intermediate structures required to predict the activities and selectivities of asymmetric catalytic reactions. Modern computational quantum chemistry has emerged as a powerful tool for explaining the selectivity and activity of asymmetric catalysts. However, reliably predicting the stereochemical outcome of realistic reactions often requires the geometry optimization of hundreds of transition state and intermediate structures, which is a tedious process. AARON automates these optimizations through an interface with a popular electronic structure package, accelerating quantum chemical workflows to enable the computational screening of potential catalysts. AARON is built using a collection of object-oriented Perl modules



Our recent work



Steve Oatley

Conclusions

CONCLUSIONS

In summary, the advantages of neural networks are:

- They are nonlinear and nonparametric; this means that few statistical assumptions are needed to apply them.
- The programs are readily available, although care and experience are necessary to use them correctly.

The corresponding disadvantages are:

- They are statistically rather poorly characterised; see Ref. 30 for a statistical critique.
- They can be slow; in a large comparative study of classification algorithms [31], back-propagation was found to be over an order of magnitude slower than simple machine learning and statistical algorithms.
- It is difficult to interpret their meaning in chemical terms; this is an active research field.

Artificial intelligence

Model/capture large amounts of data, mimic, and interpolate

Artificial intelligence

Model/capture large amounts of data, mimic, and interpolate

Human intelligence

Artificial intelligence

Model/capture large amounts of data, mimic, and interpolate

Human intelligence

Conscious of these limitations, we questioned whether an alternative catalysis manifold could be developed in which the oxidation state of phosphorus was invariant (28, 29). Such a manifold would require the unconventional step of generating a Mitsunobu-active phosphorus spe-

Artificial intelligence

Model/capture large amounts of data, mimic, and interpolate

RESEARCH

ORGANIC CHEMISTRY

Redox-neutral organocatalytic Mitsunobu reactions

Rhydian H. Beddoe¹, Keith G. Andrews¹, Valentin Magné¹, James D. Cuthbertson¹, Jan Saska¹, Andrew L. Shannon-Little¹, Stephen E. Shanahan², Helen F. Sneddon³, Ross M. Denton^{1*}

Nucleophilic substitution reactions of alcohols are among the most fundamental and strategically important transformations in organic chemistry. For over half a century, these reactions have been achieved by using stoichiometric, and often hazardous, reagents to activate the otherwise unreactive alcohols. Here, we demonstrate that a specially designed phosphine oxide promotes nucleophilic substitution reactions of primary and secondary alcohols in a redox-neutral catalysis manifold that produces water as the sole by-product. The scope of the catalytic coupling process encompasses a range of acidic pronucleophiles that allow stereospecific construction of carbon-oxygen and carbon-nitrogen bonds.

Human intelligence

Conscious of these limitations, we questioned whether an alternative catalysis manifold could be developed in which the oxidation state of phosphorus was invariant (28, 29). Such a manifold would require the unconventional step of generating a Mitsunobu-active phosphorus spe-

¹GlaxoSmithKline Carbon Neutral Laboratories for Sustainable Chemistry, School of Chemistry, University of Nottingham, Nottingham NG7 2GA, UK. ²Jealott's Hill International Research Centre, Jealott's Hill, Bracknell, Berkshire RG42 6EY, UK. ³Green Chemistry, GlaxoSmithKline, Medicines Research Centre, Stevenage, Hertfordshire SG1 2NY, UK.

*Corresponding author. Email: ross.denton@nottingham.ac.uk

Beddoe *et al.*, *Science* **365**, 910–914 (2019) 30 August

Acknowledgments

- Group members:

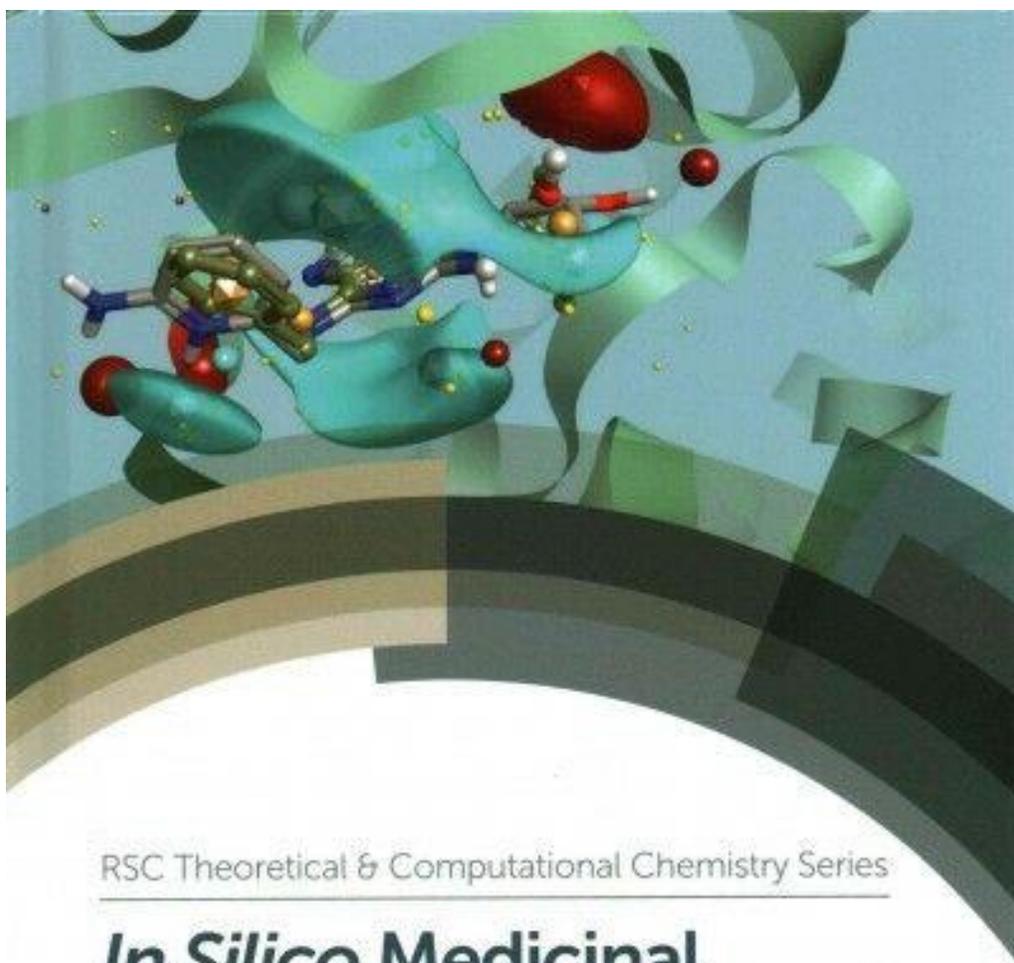
Steve Oatley, Ellen Guest, Alexe Haywood
Joe Redshaw

- Collaborators

Thomas Gaertner (Computer Science)
Stephen Pickett, Adam Taylor, Andy Mason (GSK)

- Prosperity Partnership: *Accelerated Discovery and Development of New Medicines*
GSK, Strathclyde and Nottingham



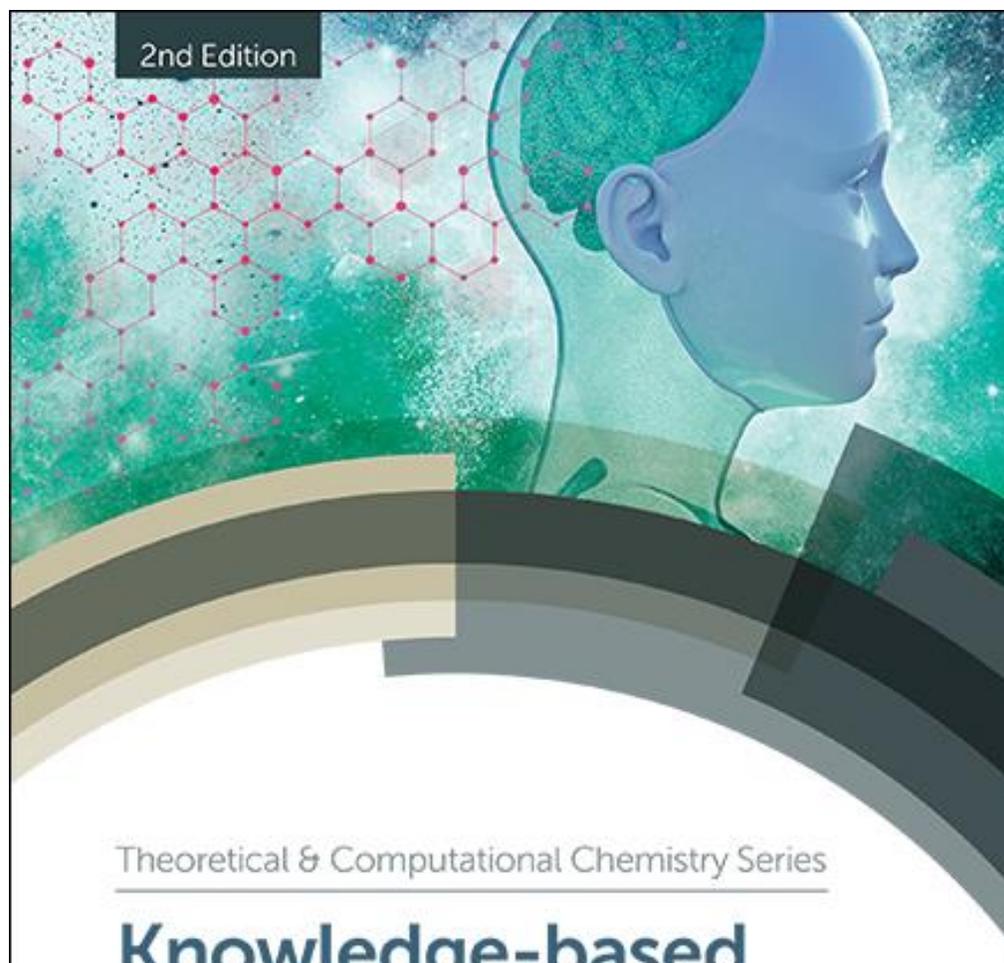


RSC Theoretical & Computational Chemistry Series

In Silico Medicinal Chemistry

Computational Methods to Support Drug Design

Nathan Brown



2nd Edition

Theoretical & Computational Chemistry Series

Knowledge-based Expert Systems in Chemistry

Artificial Intelligence in Decision Making

Philip Judson

