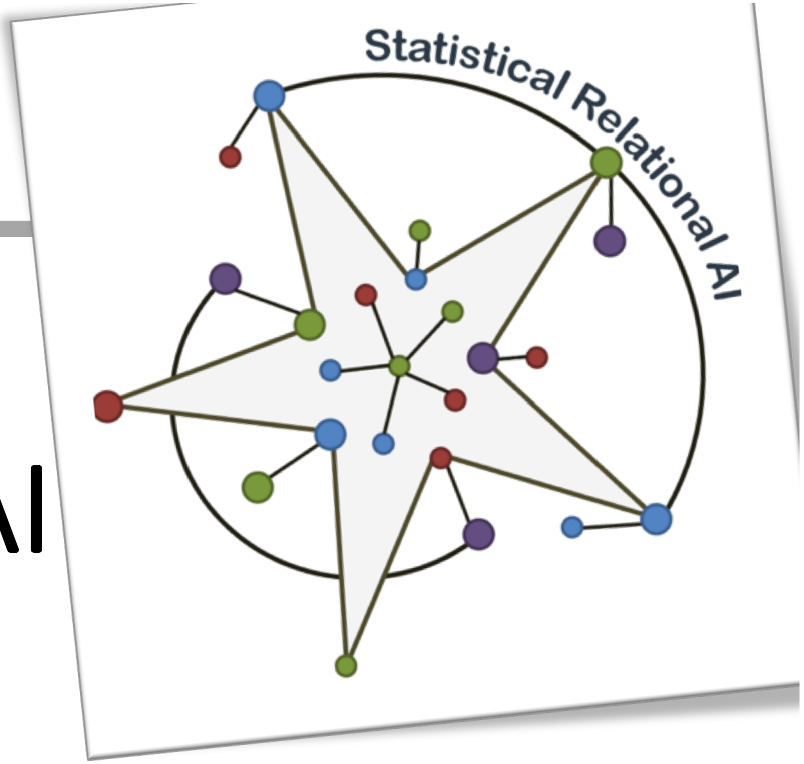
StarAI

Statistical Relational AI

Tutorial at KI-2018



Tanya Braun, Universität zu Lübeck

Kristian Kersting, Technische Universität Darmstadt

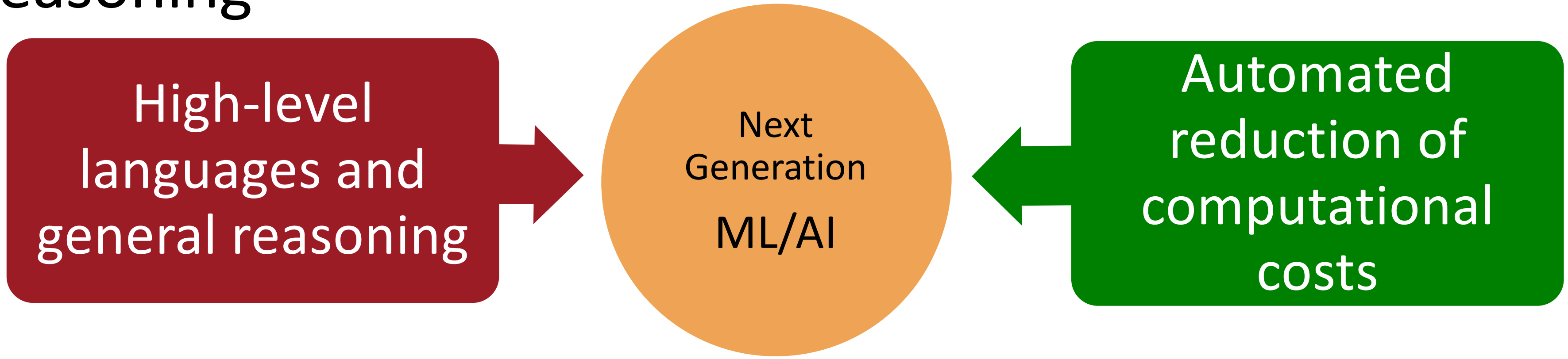
Ralf Möller, Universität zu Lübeck

UNIVERSITÄT ZU LÜBECK

TECHNISCHE UNIVERSITÄT DARMSTADT

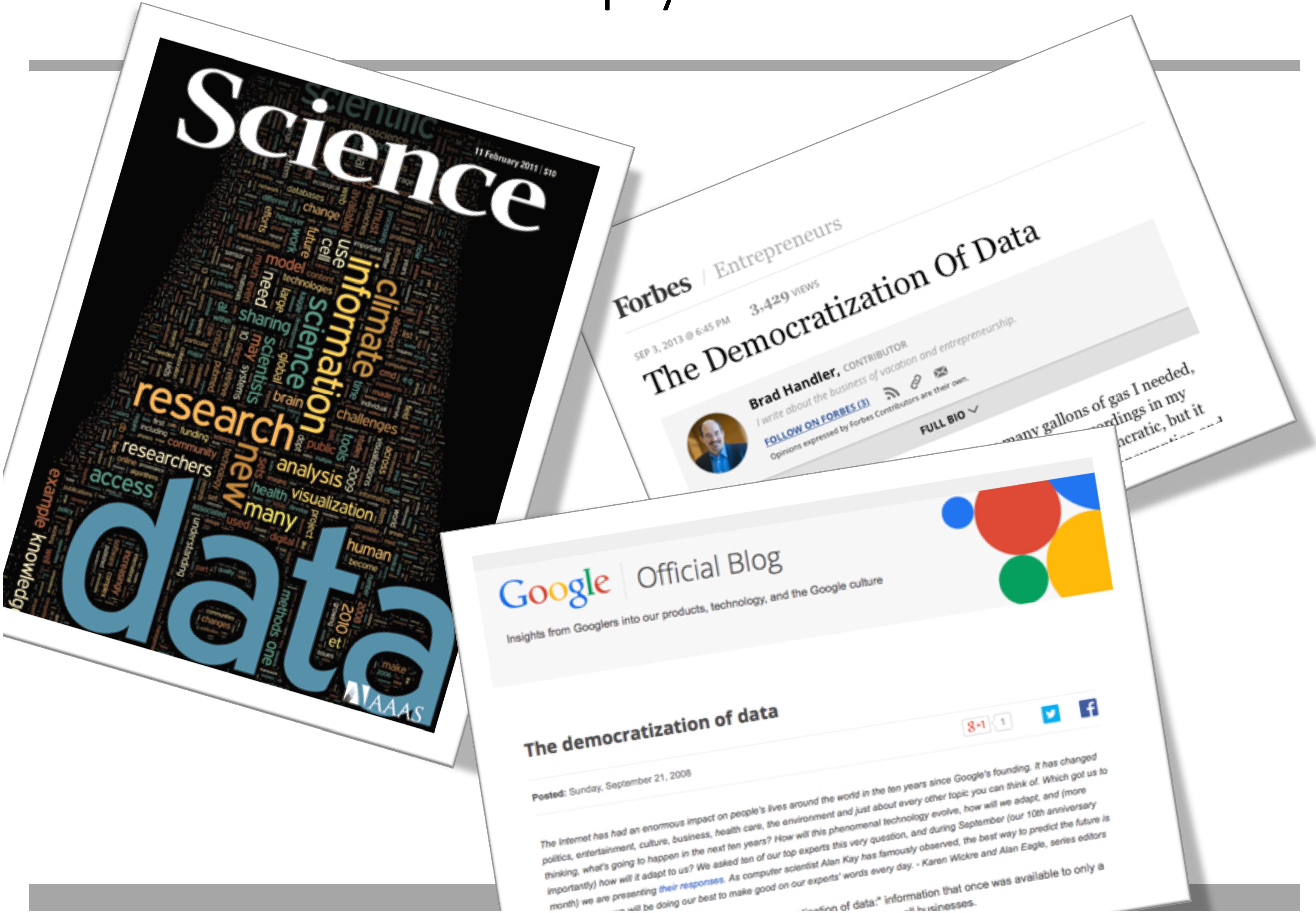
Take-away message

Statistical Machine Learning (ML) and AI need a crossover with data and programming abstractions as well as general reasoning

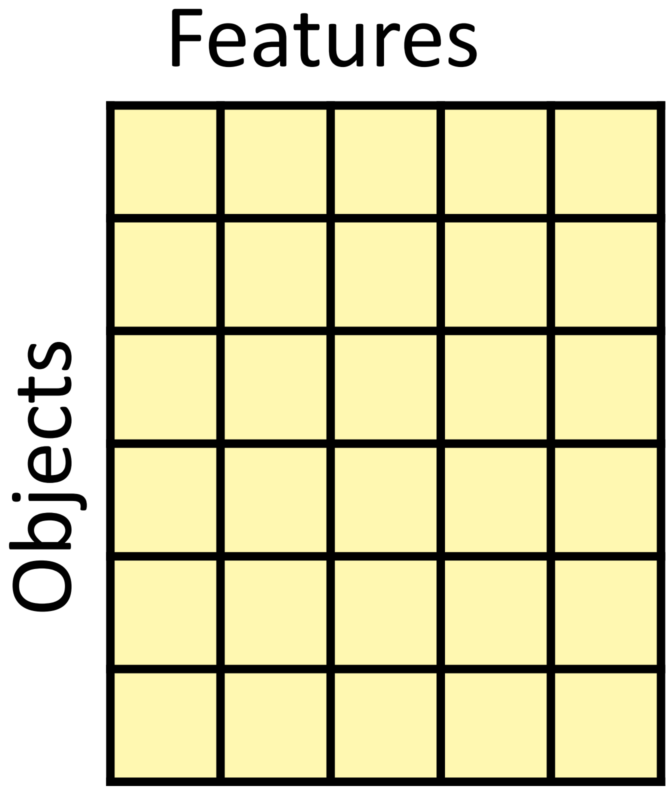


- High-level languages increase the number of people who can successfully build ML/AI applications and make experts more effective
- To deal with the computational complexity, we need ways to automatically reduce the solver costs

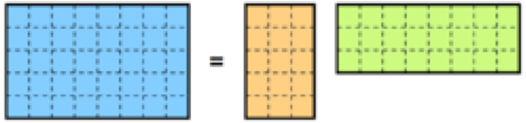
Arms race to “deeply” understand data



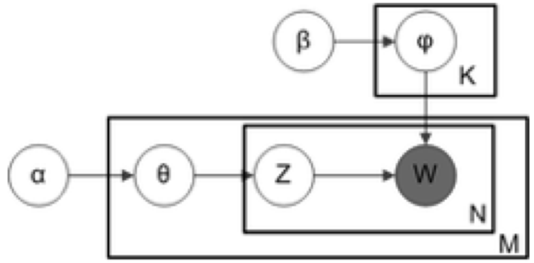
Take your spreadsheet …



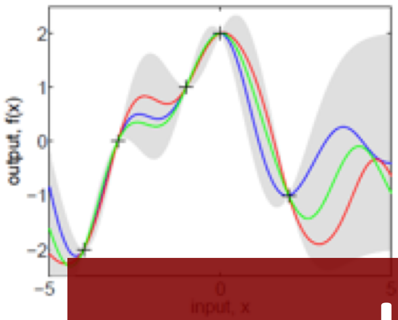
… and apply some AI/ML



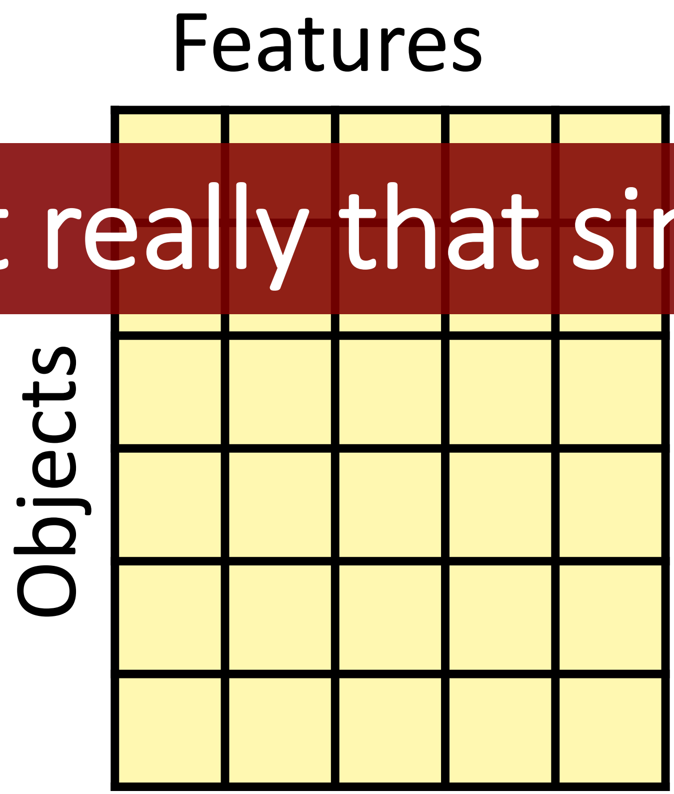
Big Data Matrix Factorization



Latent Dirichlet Allocation

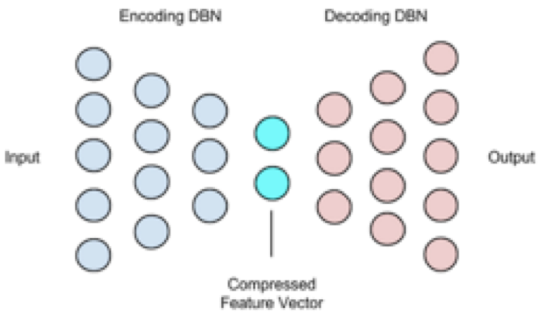


Gaussian Processes

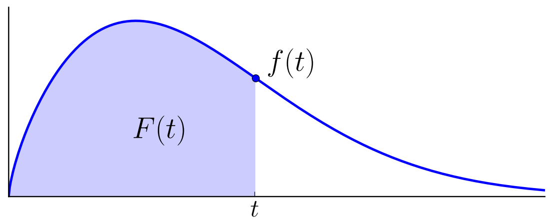


Is it really that simple?

Decision Trees/Boosting



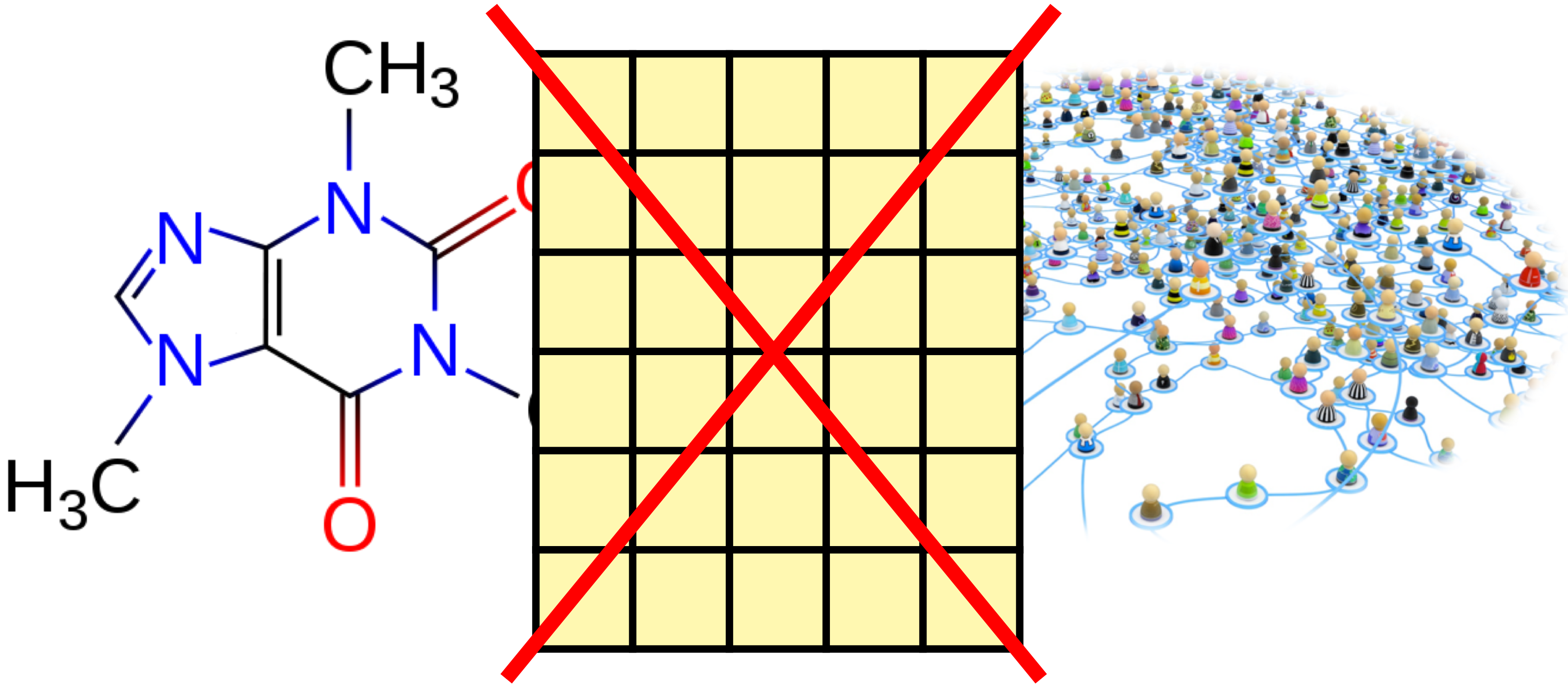
Autoencoder, Deep Learning



Diffusion Models

and many more …

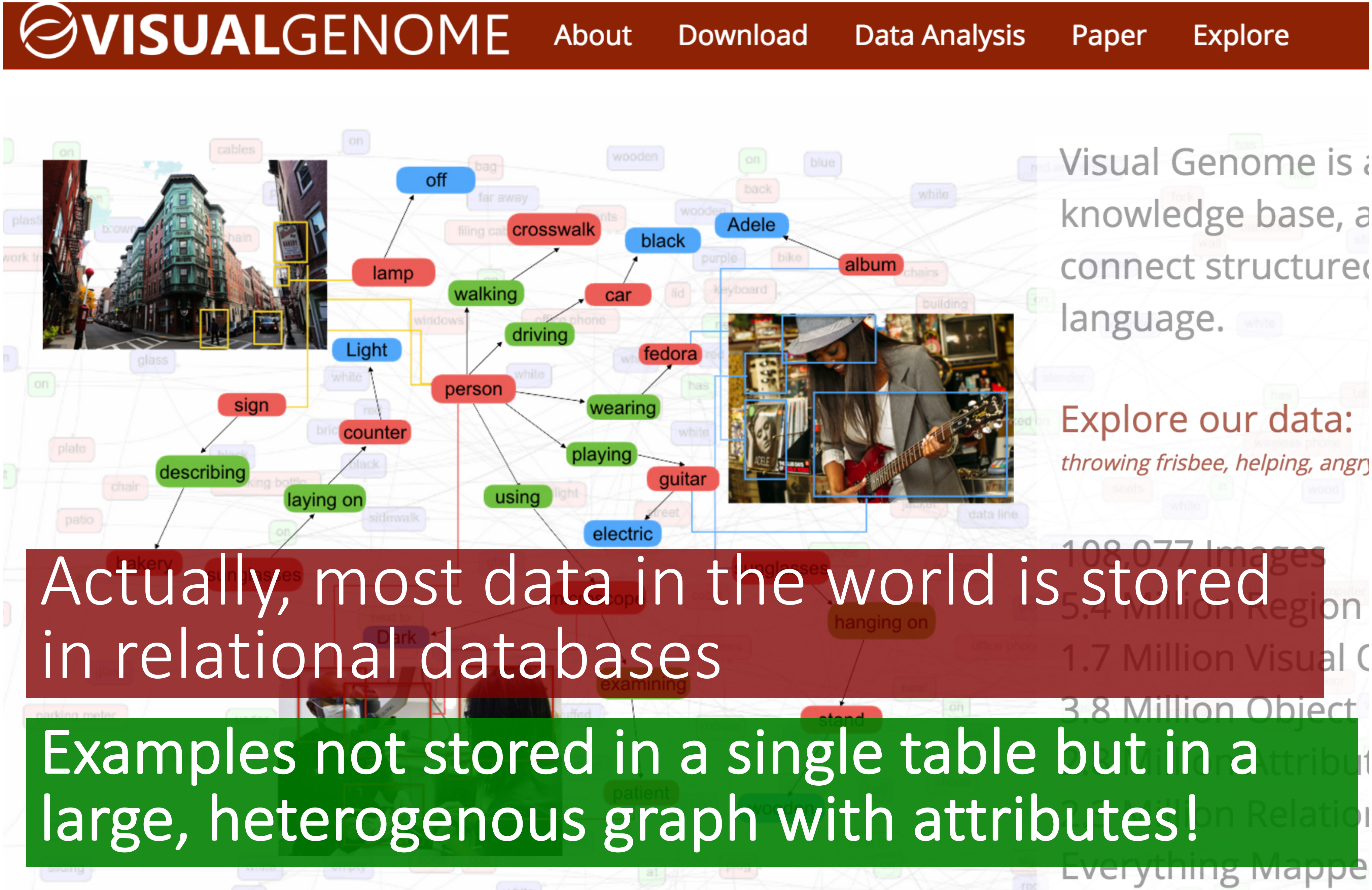
Learning and Mining with Graphs



Haussler ’99, Gärtner, Flach, Wrobel COLT’03, Vishwanathan, Schraudolph, Kondor, Borgwardt JMLR’10, Shervashidze, Schweitzer, van Leeuwen, Mehlhorn, Borgwardt JMLR’11, Neumann, Garnett, Bauckhage, Kersting MLJ’16, Morris, Kersting, Mutzel, ICDM´17, and many more

Generally, complex data networks abound

[Lu, Krishna, Bernstein, Fei-Fei „Visual Relationship Detection“ CVPR 2016]



Actually, most data in the world is stored in relational databases

Examples not stored in a single table but in a large, heterogenous graph with attributes!

Nat Rev Genet. 2012 May 2;13(6):395-405

Heart diseases and strokes – cardiovascular disease – are expensive for the world

According to the World Heart Federation, cardiovascular disease cost the European Union EURO169 billion in 2003 and the USA about EURO310.23 billion in direct and indirect annual costs. By comparison, the estimated cost of all cancers is EURO146.19 billion and HIV infections, EURO22.24 billion

Electronic Health Records A New Opportunity for AI to Save our Lifes

We have to **democratize AI, Machine Learning, and Data Science**

We have to work on **Systems AI**, so that we know how to rapidly combine, deploy, and maintain algorithms

So yes, today is the golden era of data …

… for the best-trained, best-funded Machine Learning and Artificial Intelligence teams

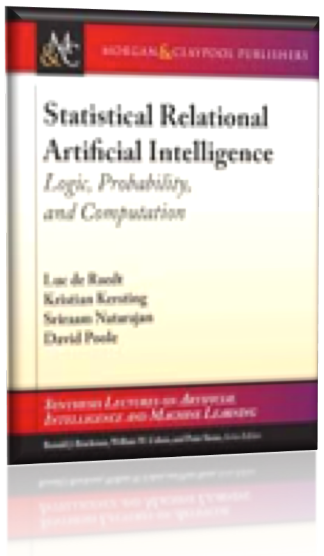
Systems AI: the computational and mathematical modeling of complex AI systems.

Eric Schmidt, Executive Chairman, Alphabet Inc.: Just Say "Yes", Stanford Graduate School of Business, May 2, 2017.https://www.youtube.com/watch?v=vbb-AjiXyh0.

Kordjamshidi, Roth, Kersting: "Systems AI: A Declarative Learning Based Programming Perspective." IJCAI-ECAI 2018.

For Systems AI we have to deeply understand data, knowledge and reasoning in a large number of forms

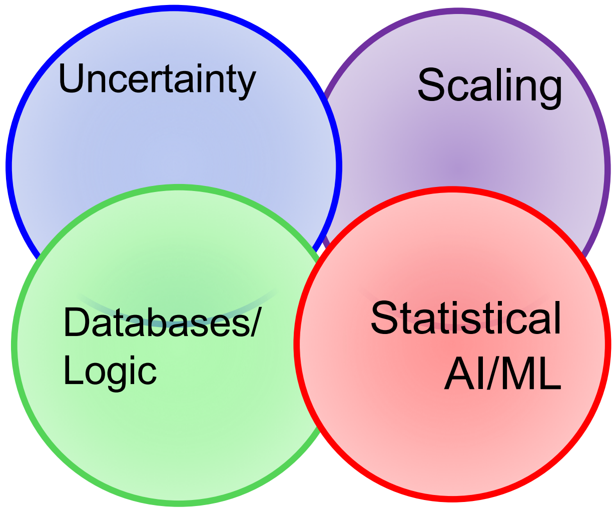
Crossover of Statistical AI/ML with data & programming abstractions



building general-purpose thinking and learning machines

make the AI/ML expert more effective

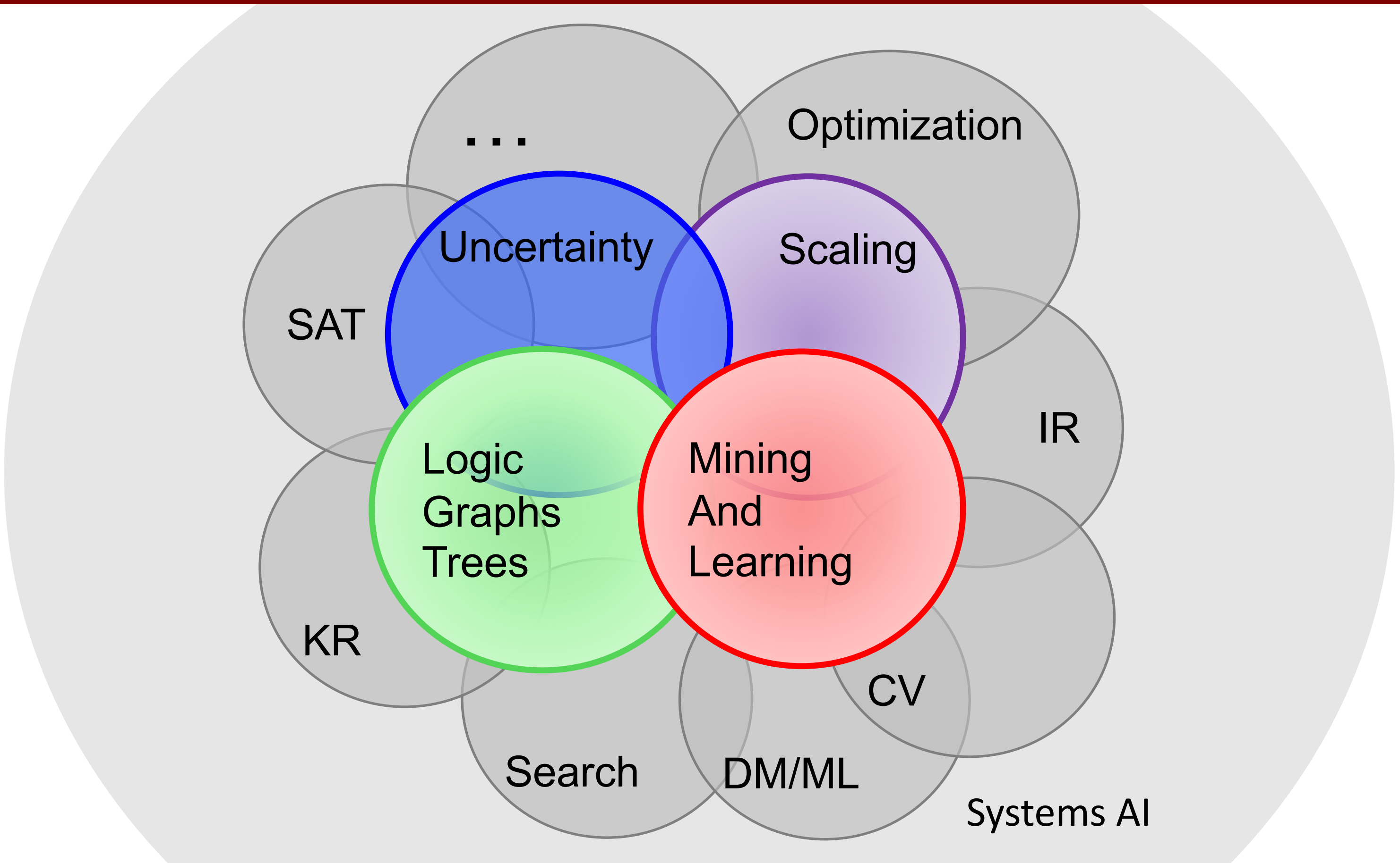
increases the number of people who can successfully build AI/ML applications



De Raedt, Kersting, Natarajan, Poole: Statistical Relational Artificial Intelligence: Logic, Probability, and Computation. Morgan and Claypool Publishers, ISBN: 9781627058414, 2016.

Statistical Relational Learning/AI

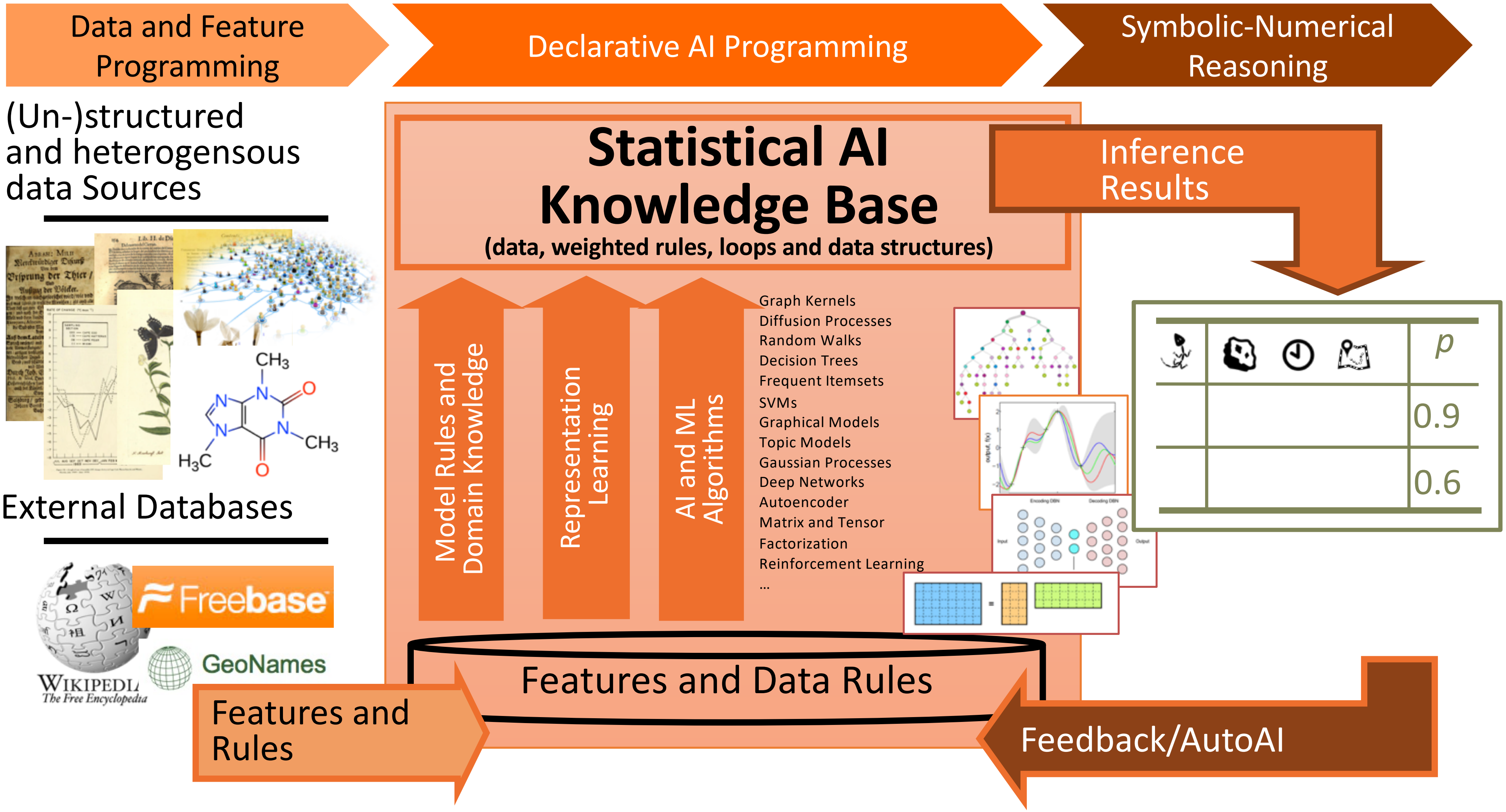
… study and design intelligent agents that reason about and act in noisy worlds composed of objects and relations among the objects

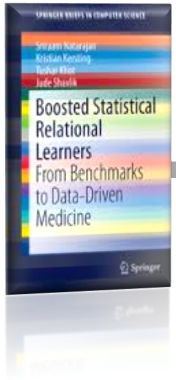


[Getoor, Taskar MIT Press '07; De Raedt, Frasconi, Kersting, Muggleton, LNCS'08; Domingos, Lowd Morgan Claypool '09; Natarajan, Kersting, Khot, Shavlik Springer Brief'15; Russell CACM 58(7): 88-97 '15, Gogate, Domingos CACM 59(7):107-115 ´16]

[Ré, Sadeghian, Shan, Shin, Wang, Wu, Zhang IEEE Data Eng. Bull.'14; Natarajan, Picado, Khot, Kersting, Ré, Shavlik ILP'14; Natarajan, Soni, Wazalwar, Viswanathan, Kersting Solving Large Scale Learning Tasks'16, Mladenov, Heinrich, Kleinhans, Gonsior, Kersting DeLBP'16, Kordjamshidi, Roth, Kersting IJCAI-ECAI 2018, ...]

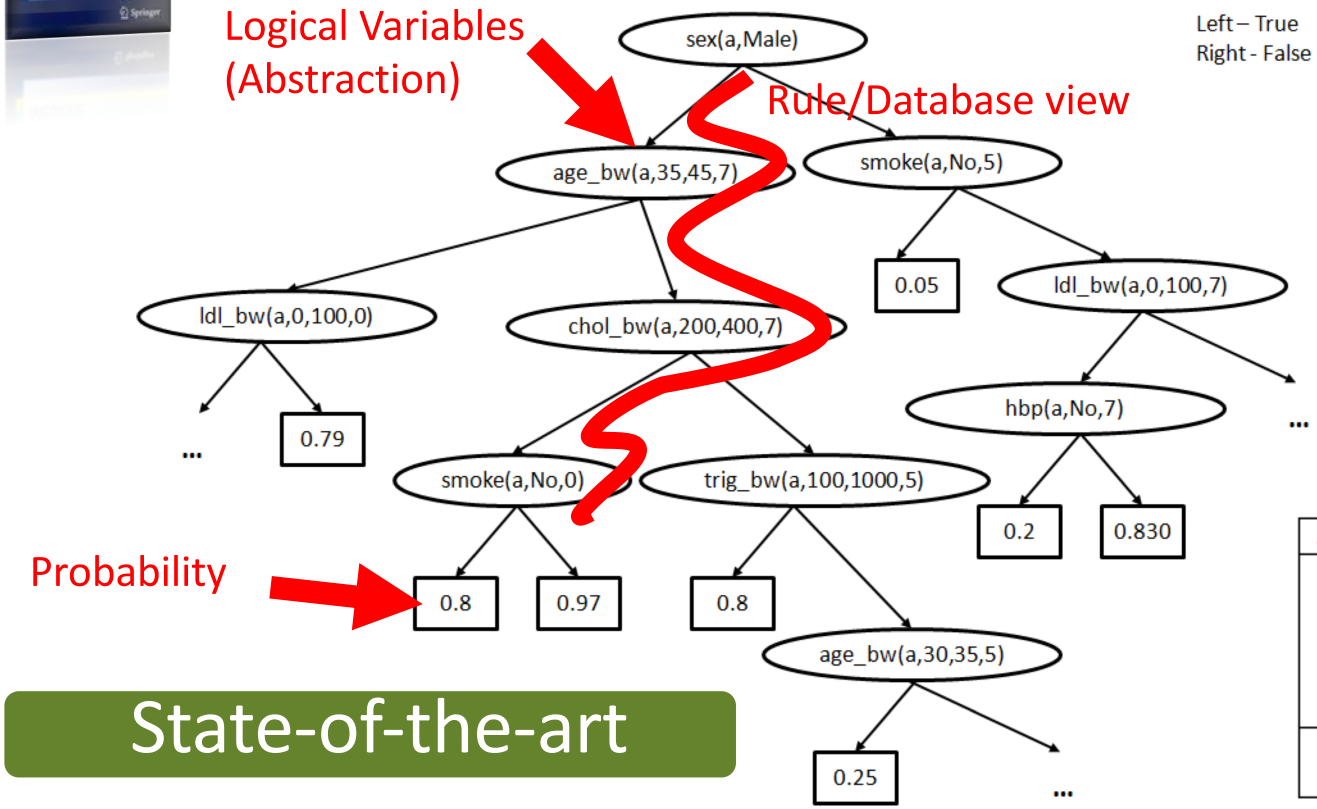
This establishes a novel "Deep AI"

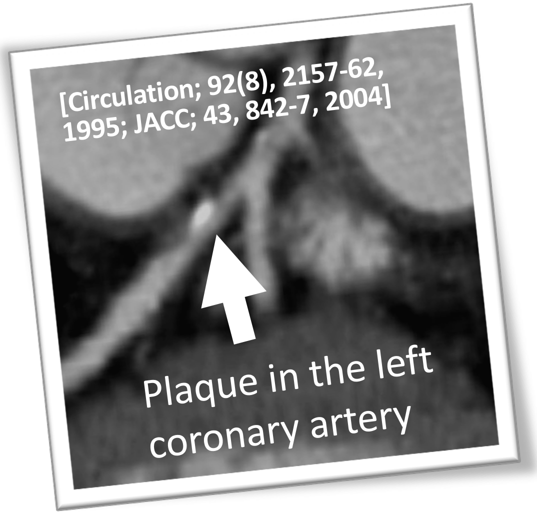




This “Deep AI” can understand EHRs

Atherosclerosis is the cause of the majority of Acute Myocardial Infarctions (heart attacks)





Algorithm	Accuracy	AUC-ROC
J48	0.667	0.607
SVM	0.667	0.5
AdaBoost	0.667	0.608
Bagging	0.677	0.613
NB	0.75	0.653
RPT	0.669*	0.778
RFGB	0.667*	0.819

The higher, the better

25%

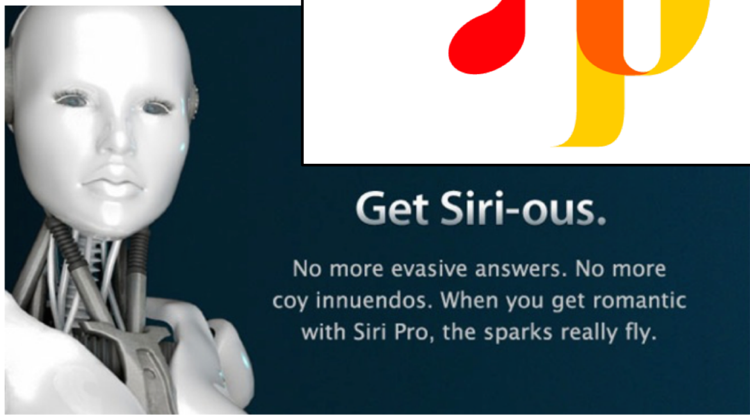
State-of-the-art

Algorithm for Mining Markov Logic Networks	Likelihood The higher, the better	AUC-ROC The higher, the better	AUC-PR The higher, the better	Time The lower, the better
Boosting	0.81	0.96	0.93	**9s**
LSM	0.73	0.54	0.62	**93 hrs**
	11%	78%	50%	37200x faster

[Kersting, Driessens ICML´08; Karwath, Kersting, Landwehr ICDM´08; Natarajan, Joshi, Tadepelli, Kersting, Shavlik. IJCAI´11; Natarajan, Kersting, Ip, Jacobs, Carr IAAI `13; Yang, Kersting, Terry, Carr, Natarajan AIME ´15; Khot, Natarajan, Kersting, Shavlik ICDM´13, MLJ´12, MLJ´15]

This „Deep AI“ excites industry

RelationalAI, Infor, Apple, and Uber are investing hundreds of millions of US dollars



IBM

CPLEX

GUROBI OPTIMIZATION

And it appears in industrial strength solvers such as CPLEX and GUROBI

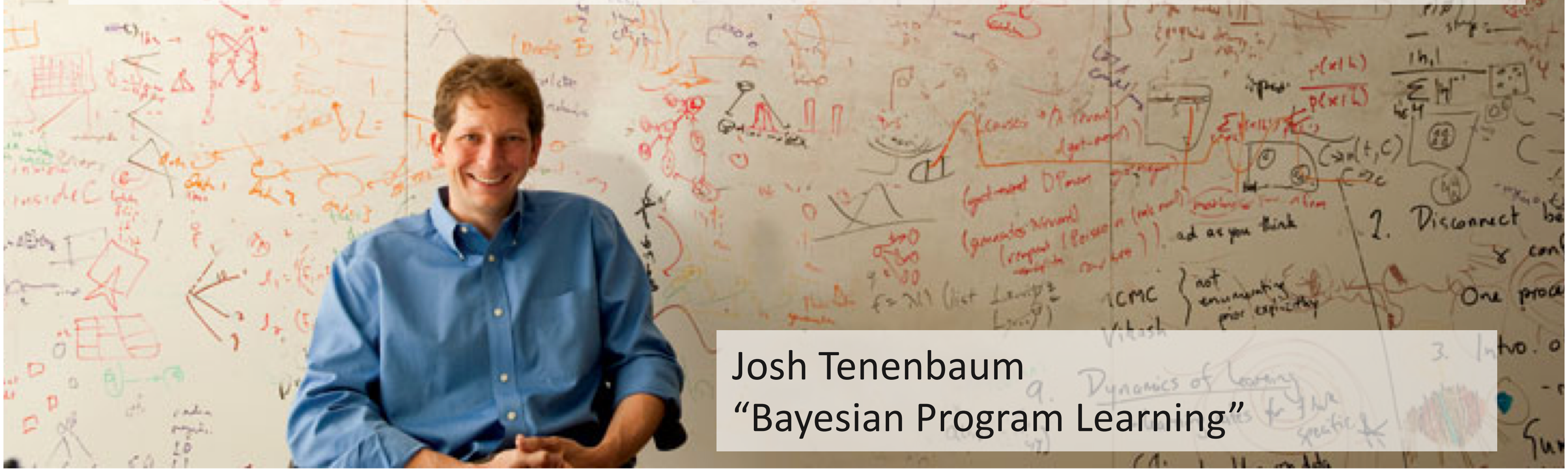
This „Deep AI“ connects well to DB theory

Jim Gray Turing Award 1998
“Automated Programming”

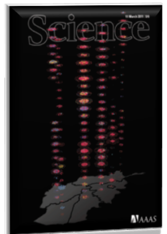
Mike Stonebraker Turing Award 2014
“One size does not fit all”

… and cognitive science

"How do we humans get so much from so little?" and by that I mean how do we acquire our understanding of the world given what is clearly by today's engineering standards so little data, so little time, and so little energy.

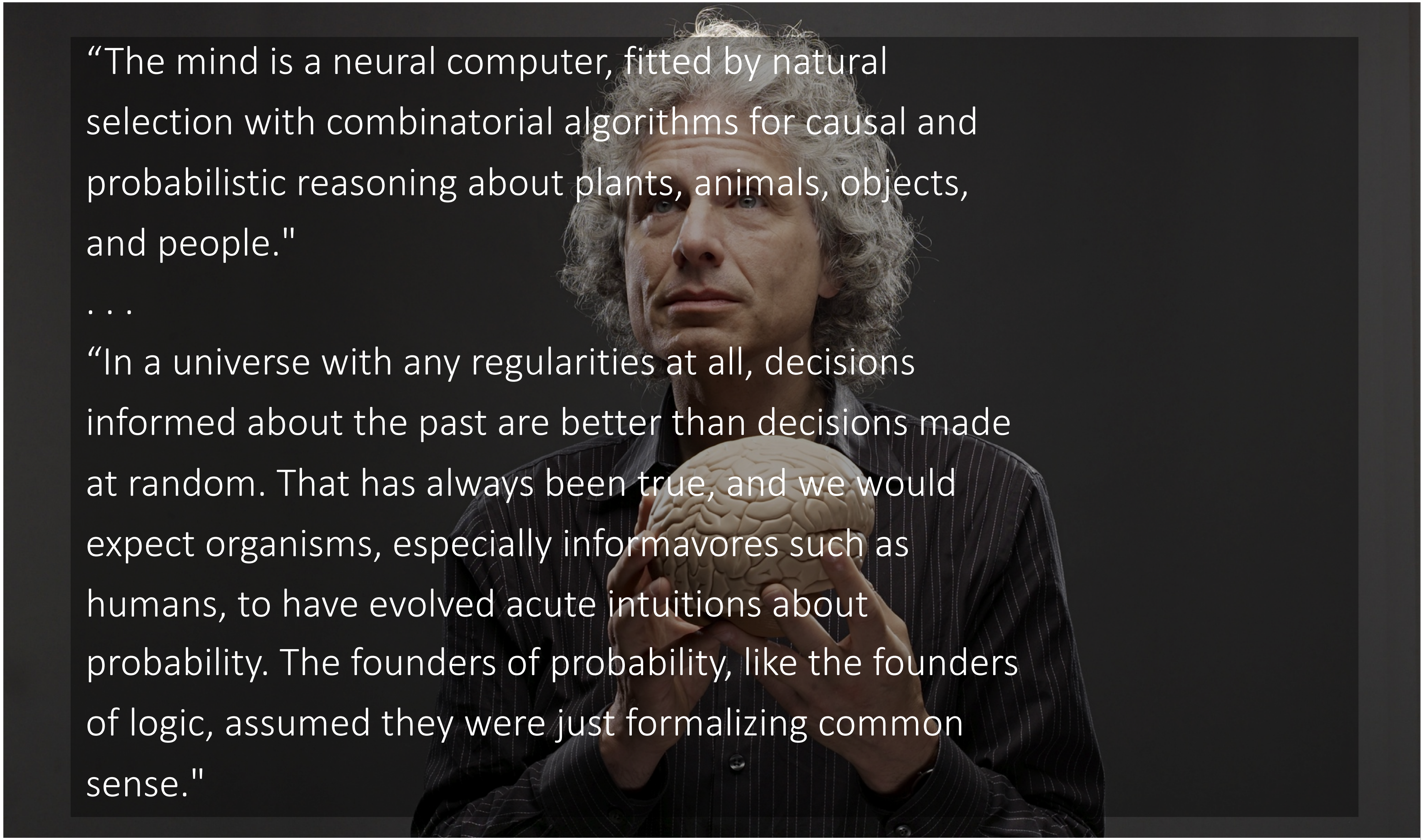


Josh Tenenbaum
"Bayesian Program Learning"



Lake, Salakhutdinov, Tenenbaum, Science 350 (6266), 1332-1338, 2015
Tenenbaum, Kemp, Griffiths, Goodman, Science 331 (6022), 1279-1285, 2011

… and it is indeed deep



“The mind is a neural computer, fitted by natural selection with combinatorial algorithms for causal and probabilistic reasoning about plants, animals, objects, and people."

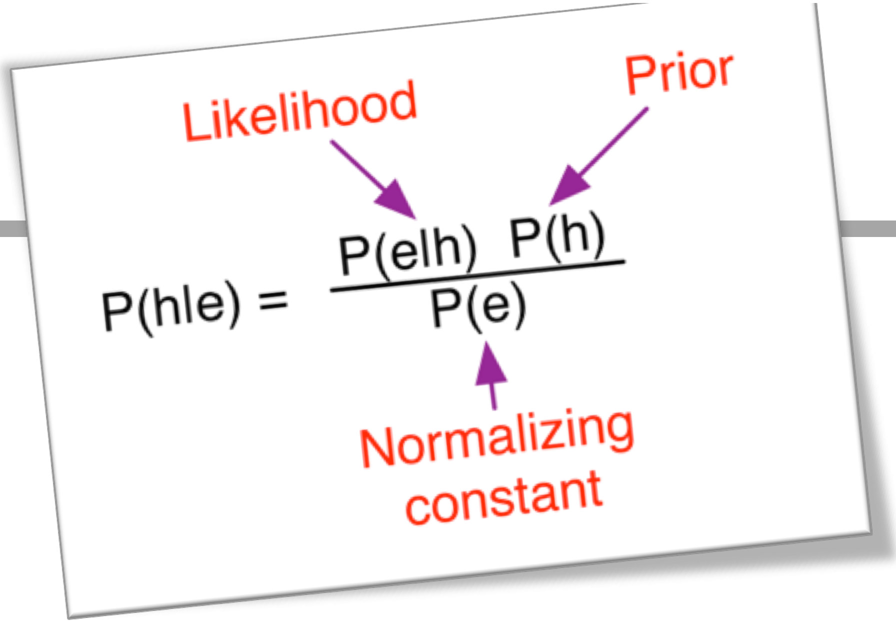
. . .

“In a universe with any regularities at all, decisions informed about the past are better than decisions made at random. That has always been true, and we would expect organisms, especially informavores such as humans, to have evolved acute intuitions about probability. The founders of probability, like the founders of logic, assumed they were just formalizing common sense."

-Steven Pinker, How the Mind Works, 1997, pp. 524, 343.

Let’s consider some more gentle examples

Bayes' Rule



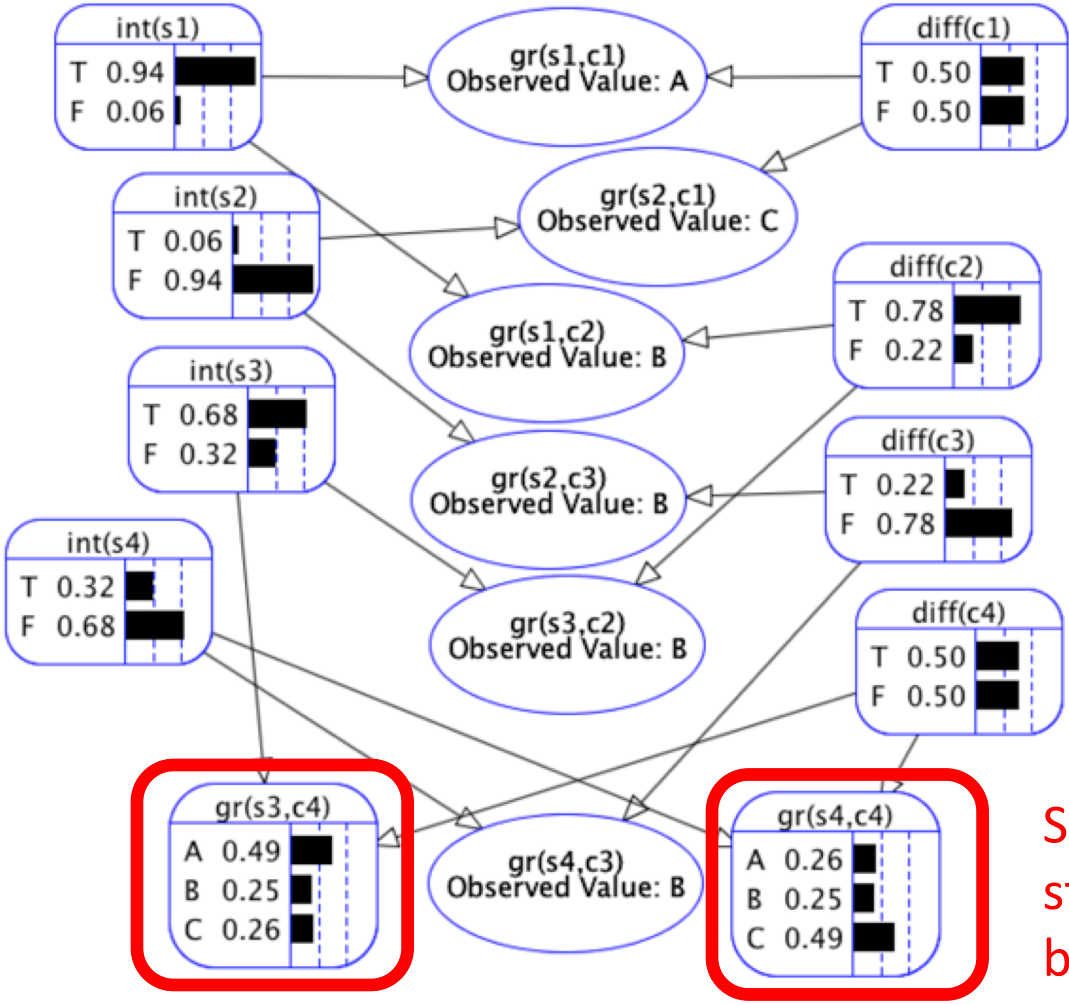
- What if h is the effect of a drug on a particular patient, and e is the patient's electronic health record?
- What if e is the electronic health records for all of the people in the world?
- What if e is a collection of student records in a university?
- What if e is a description of everything known about the geology of Earth?

Predicting Predicting Grades

- Students s3 and s4 have the same averages, on courses with the same averages.
- Which student would you expect to do better?

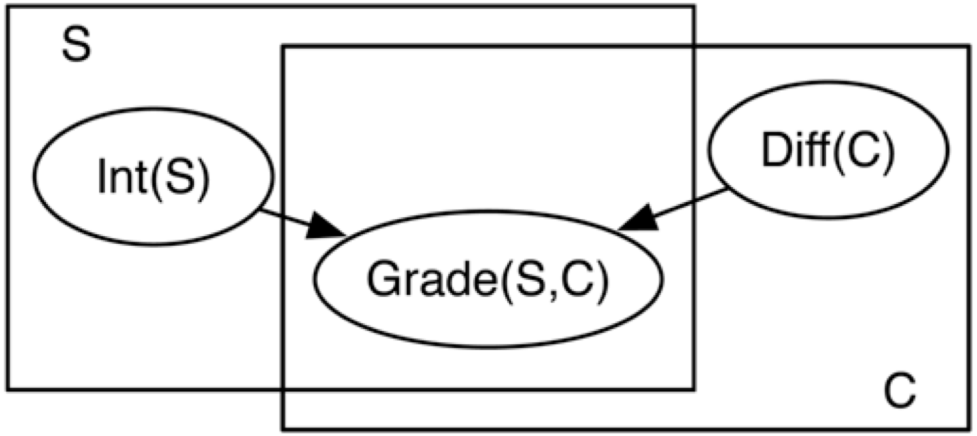
Student	Course	Grade
s_1	c_1	A
s_2	c_1	C
s_1	c_2	B
s_2	c_3	B
s_3	c_2	B
s_4	c_3	B
s_3	c_4	?
s_4	c_4	?

Rigid and Large Graphical Model for Predicting Grades



So, we should expect student s3 to perform better

A more flexible and compact way of predicting grades: Relational Models



Using plate notation, one can captures the regularities

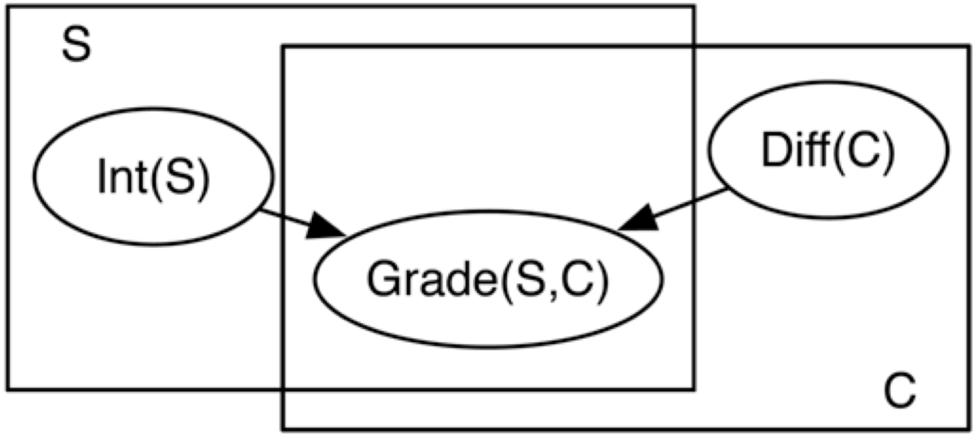
Program Abstraction:

- S, C logical variable representing students, courses
- the set of individuals of a type is called a population
- Int(S), Grade(S, C), D(C) are parametrized random variables

Grounding:

- for every student s, there is a random variable Int(s)
- for every course c, there is a random variable Di(c)
- for every s, c pair there is a random variable Grade(s,c)
- all instances share the same structure and parameters

A more flexible and compact way of predicting grades: Relational Models

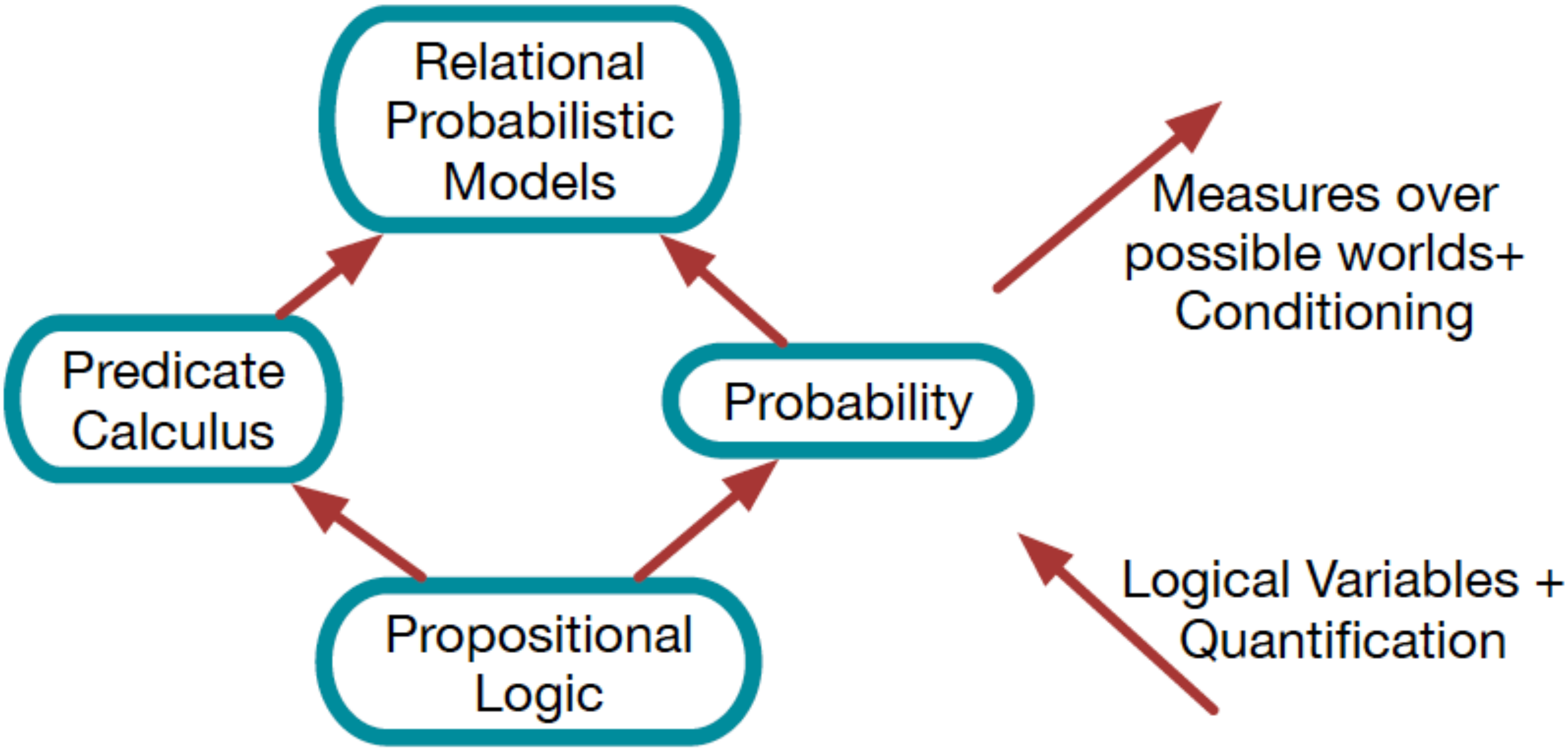


Using plate notation, one can captures the regularities

- If there were 1000 students and 100 courses:
  - Grounding contains
    - 1000 I(s) variables
    - 100 D(c) variables
    - 100000 Gr(s,c) variables
  - total: 101100 variables
- Numbers to be specified to define the probabilities:

  1 for I (S), 1 for D(C), 8 for Gr(S,C) = 10 parameters.

Relational Probabilistic Models



Relational Probabilistic Models

Random variables for combinations of individuals in populations

- build a probabilistic model before knowing (all of) the individuals
- learn the model for one set of individuals
- apply the model to existing and new individuals
- allow complex relationships between individuals

Exchangeability:

- Before we know anything about individuals, they are indistinguishable, and so should be treated identically.

Uncertainty about:

- Properties of individuals
- Relationships among individuals
- Identity (equality) of individuals
- Existence (and number) if individuals

Mission and Schedule of the Tutorial*

Providing an overview and a synthesis of StarAI

- Introduction (Kristian) — 10 min ✓
  - Star AI, Systems AI
- Overview: Probabilistic relational modeling (Ralf) — 40 min
  - Semantics (grounded-distributional, maximum entropy)
  - Inference problems and their applications
  - Algorithms and systems
  - Scalability (limited expressivity, knowledge compilation, approximation)
- Scalability by lifting
  - Exact lifted inference (Tanya) — 40+90 min
  - Approximate lifted inference (Kristian) — 10 min
- Learning (Kristian) — 15 min
  - Parameter learning (stochastic gradient descent)
  - Structure learning
  - Relational reinforcement learning
- Summary — 5 min

*We thank the SRL/StarAI crowd for all their exciting contributions! The tutorial is necessarily incomplete and we apologize to anyone whose work we are not citing