Andreas Holzinger

706.315 Selected Topics on Knowledge Discovery: Interactive Machine Learning

2015W, SE, 2.0 h, 3.0 ECTS

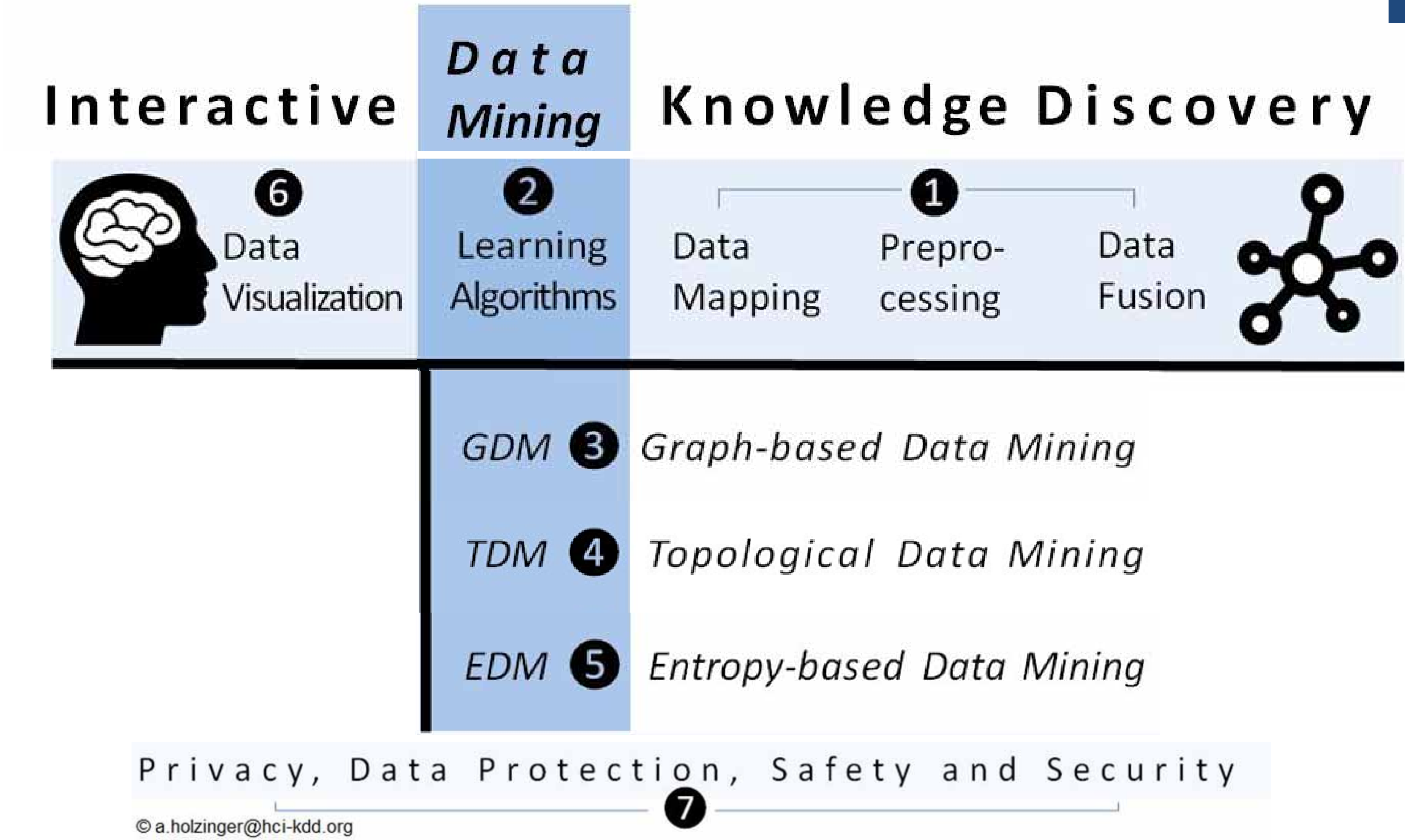
Week 44 - 30.10.2015 10:00-11:30

2 - Human Learning & Machine Learning

a.holzinger@hci-kdd.org

http://hci-kdd.org/lv-706-315-interactive-machine-learning

In the Knowledge Discovery pipeline ML is the heart …



Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: Cognitive Science meets Machine Learning. IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

Agenda

- 01 Why is Cognitive Science important for ML?
- 02 When are humans better than computers?
- 03 On Human Information Processing
- 04 Decision Making under Uncertainty
- 05 Graphical Models and Decision Making
- 06 Probabilistic Programming
- 07 Conclusion
- 08 Questions
- 09 Appendix

Quiz: In which tasks are humans better than computers?


1

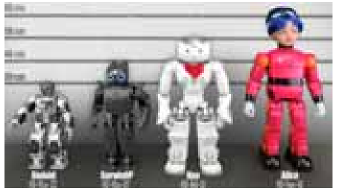

2


3

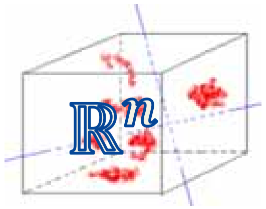

4


5

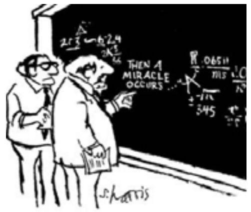
6


7

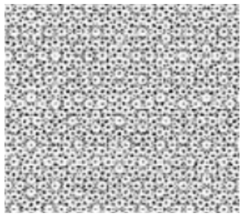

8

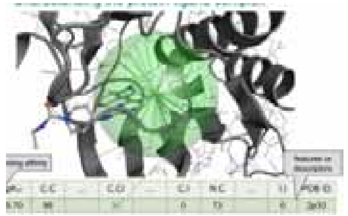


9



"I think you should be more explicit here in step two."
10


11


12

01 Why is Cognitive Science important for ML ?

- **Cognitive Science → human intelligence**
  - Study the principles of ***human learning*** to understand biological intelligence
- **Human-Computer Interaction → the bridge**
  - Interacting with algorithms that learn shall enhance user friendliness and let concentrate on problem solving - Opening the “black-box” to a “glass-box”
- **Computer Science → computational intelligence**
  - Study the principles of ***machine learning*** to understand artificial intelligence

What is this: Cognitive Science?

- "By 1960 it was clear that something interdisciplinary was happening. At Harvard we called it cognitive studies, at Carnegie-Mellon they called it information-processing psychology, and at La Jolla they called it cognitive science. "

George A. Miller (1920-2012), Harvard University, well known for:

The magical number seven, plus or minus two: Some limits on our capacity for processing information.
GA **Miller** - Psychological review, 1956 - psycnet.apa.org
Abstract 1. A variety of researches are examined from the standpoint of information theory. It is shown that the unaided observer is severely limited in terms of the amount of information he can receive, process, and remember. However, it is shown that by the use of various ...
Zitiert von: 23560 Ähnliche Artikel Alle 70 Versionen Web of Science: 7697 In EndNote importierer

- CS aims to reverse engineer **human intelligence**;
- ML provides powerful sources of insight into ***how*** **machine intelligence** is possible.
- CS therefore raises challenges for, and draws inspiration from ML;
- ML could inspire **new directions** by novel insights about the human mind

Some definitions in Cognitive Science (very incomplete)

- Intelligence
  - Hundreds of controversial definitions – very hard to define; related terms include the ability to solve problems, make decisions and acquire and apply knowledge and skills.
- Learning
  - Different definitions – basically acquisition of knowledge through experience, study or being taught
- Problem Solving
  - Process of finding solutions to complex issues
- Reasoning
  - ability of our mind to think and understand things
- Decision Making
  - Process of “de-ciding” (“ent-scheiden”) between alternative options
- Sense Making
  - Process of giving meaning to experience
- Causality
  - Relationship between cause and effect

Typical Questions of Cognitive Science Research

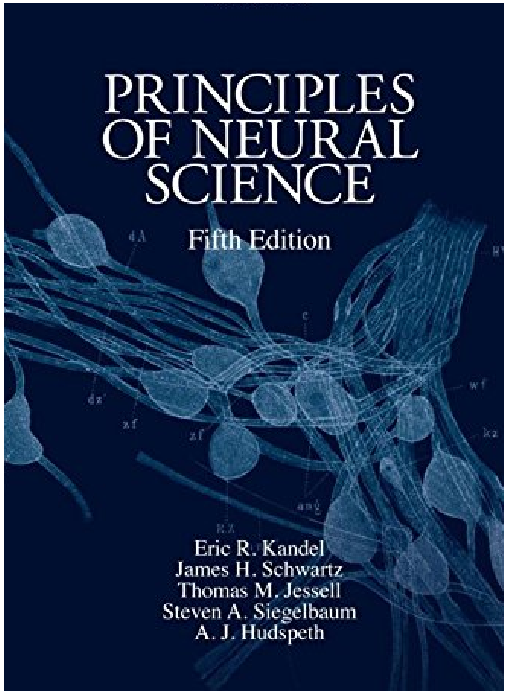
- How does our mind work?
- How do we process information?
- How do we learn and generalize?
- How do we solve complex problems?
- How do we reason and make decisions?
- How do we make predictions?
- How do we behave in new situations?
- How can we build intelligent agents?

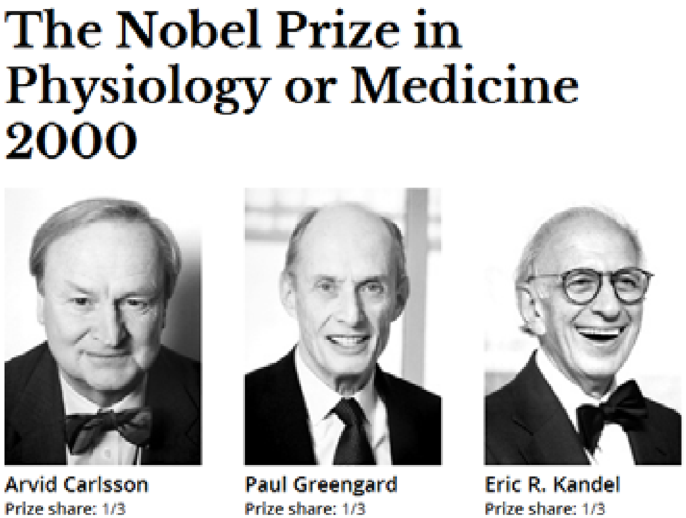
“Solve intelligence – then solve everything else”

Demis Hassabis, 22 May 2015

The Royal Society,
Future Directions of Machine Learning Part 2

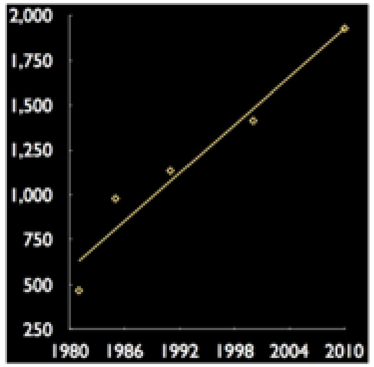
https://youtu.be/XAbLn66iHcQ?t=1h28m54s





This book doubled in Volume every decade …

Number of Pages ->



Editions ->

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A. & Hudspeth, A. 2012. Principles of neural science, 5th Edition (1760 pages), New York: McGraw-Hill.

- Facts ≠ Knowledge, Descriptions ≠ Insight
- **Our goal should be the opposite: To make this book shorter!**

CS vs ML did NOT harmonize in the past

- CS had its focus on specific experimental paradigms because it was embedded deeply in Psychology and Linguistics; and aimed to be cognitively/neutrally plausible …
- ML had its focus on standard learning problems and tried to optimize in the range of 1 % because it was embedded in Computer Engineering; and aimed to have working systems whether mimicking the human brain or not …

Wilhelm Wundt (1832 – 1920): Psychology as Science

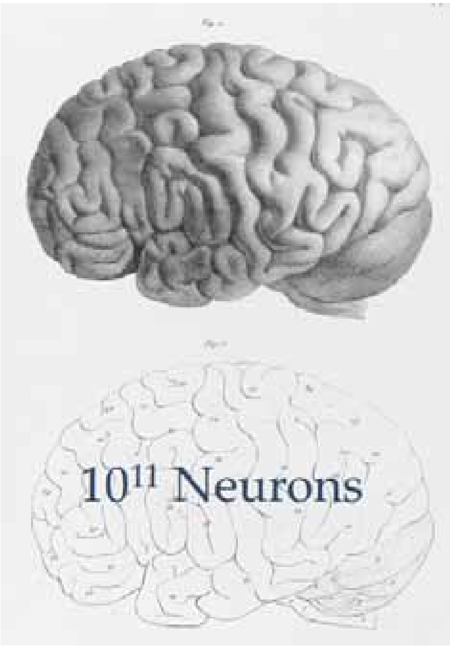
- First experimental psychology laboratory at Leipzig, in 1879
- **Structuralism**: "Human mental experience, no matter how complex, can be viewed as blends or combinations of simple processes or elements."
- Influenced by John Stuart Mill's – **mental chemistry**.
- rather than **computational components**, building blocks are **subjective experience** (qualia)

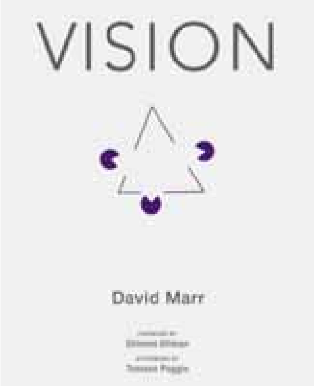
Pioneers of Psychology

Wilhelm Wundt (1832–1920). *[Archives of the History of American Psychology, University of Akron].*

David Marr (1945 – 1980) Neuroscientist

- Cerebellum: big memory to support motor learning
- Neocortex: big memory flexibly learns statistical structure from input patterns
- Hippocampus: big memory encoding memory traces via Hebbian learning
- Example Vision: process of discovering properties (what, where) of things in the real-world from 3D-images
- Vision = information processing task + rich internal representation
- Understanding of vision requires multiple levels of analysis: computational – algorithmic and implementational

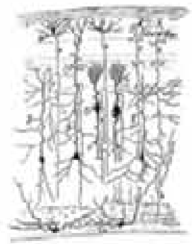




- **Computation**
  - "What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?"

$$V := \sup_{(\tau,\mu)} \frac{\mathbb{E}\left[1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^{m} r_j 1_{\{\mu=j, M=j\}}\right]}{\mathbb{E}\left[(\tau+T_0)\wedge\Theta\right]}$$

- **Representation and algorithm**
  - "What is the representation for the input and output, and the algorithm for the transformation?"



- **Implementation**
  - "How can the representation and algorithm be realized physically?"

Cognitive Science versus Machine Learning

- Human learning
- Categorization
- Causal learning
- Function learning
- Representations
- Language
- Experiment design

- Machine learning
- Density estimation
- Graphical models
- Regression
- Nonparametric Bayes
- Probabilistic grammars
- Inference algorithms

We pick one example …

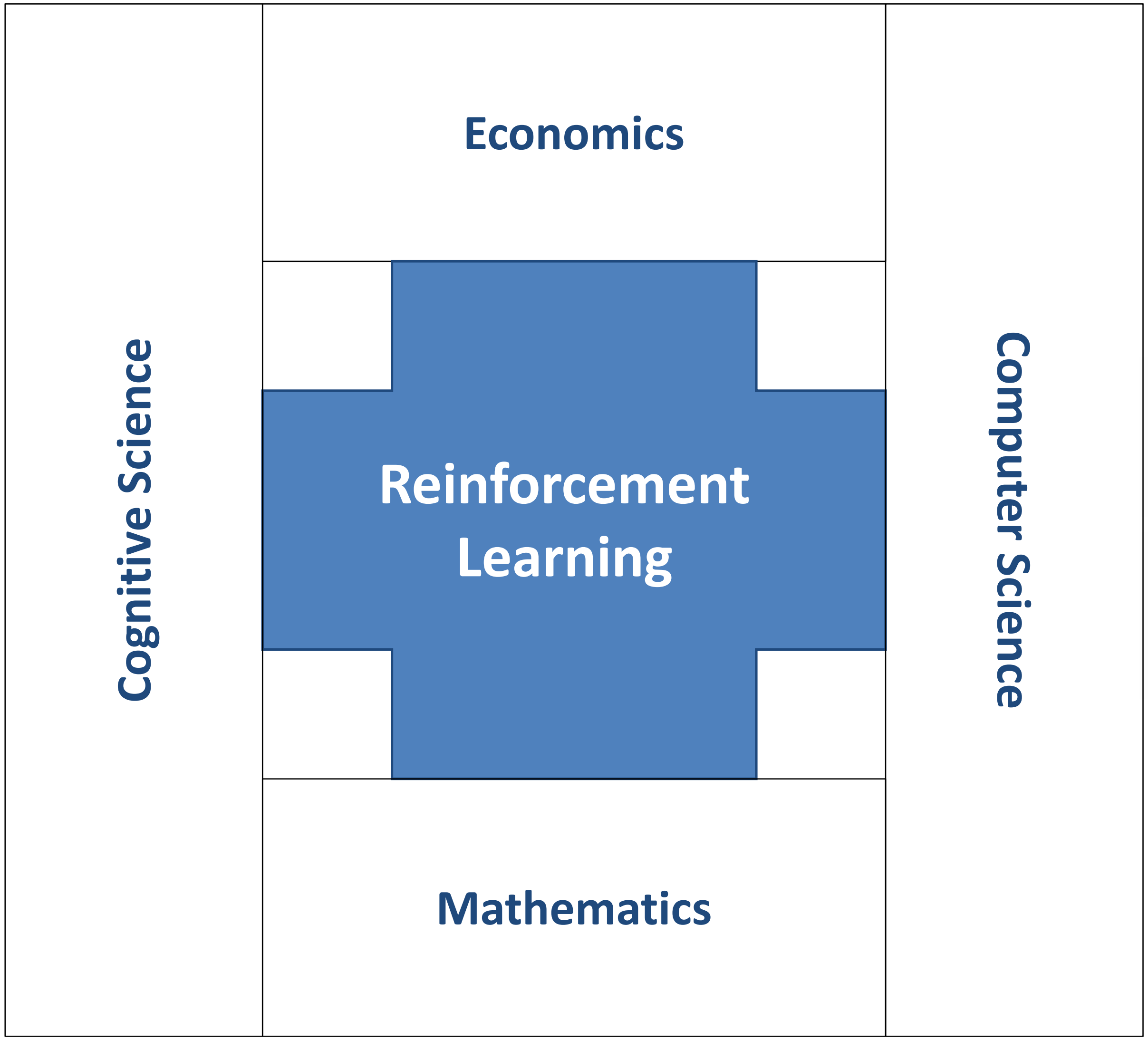
- **Causal learning** - how can we get insights from studying CL and what might make better ML systems from studying causal learning

Is the human brain a inference engine ?

- Learning concepts from examples (babies!)
- Causal inference and reasoning
- Predicting everyday events
- Even little children solve complex problems unconsciously, effortlessly, and … successfully
- Compare your best Machine Learning algorithm with a seven year old child!

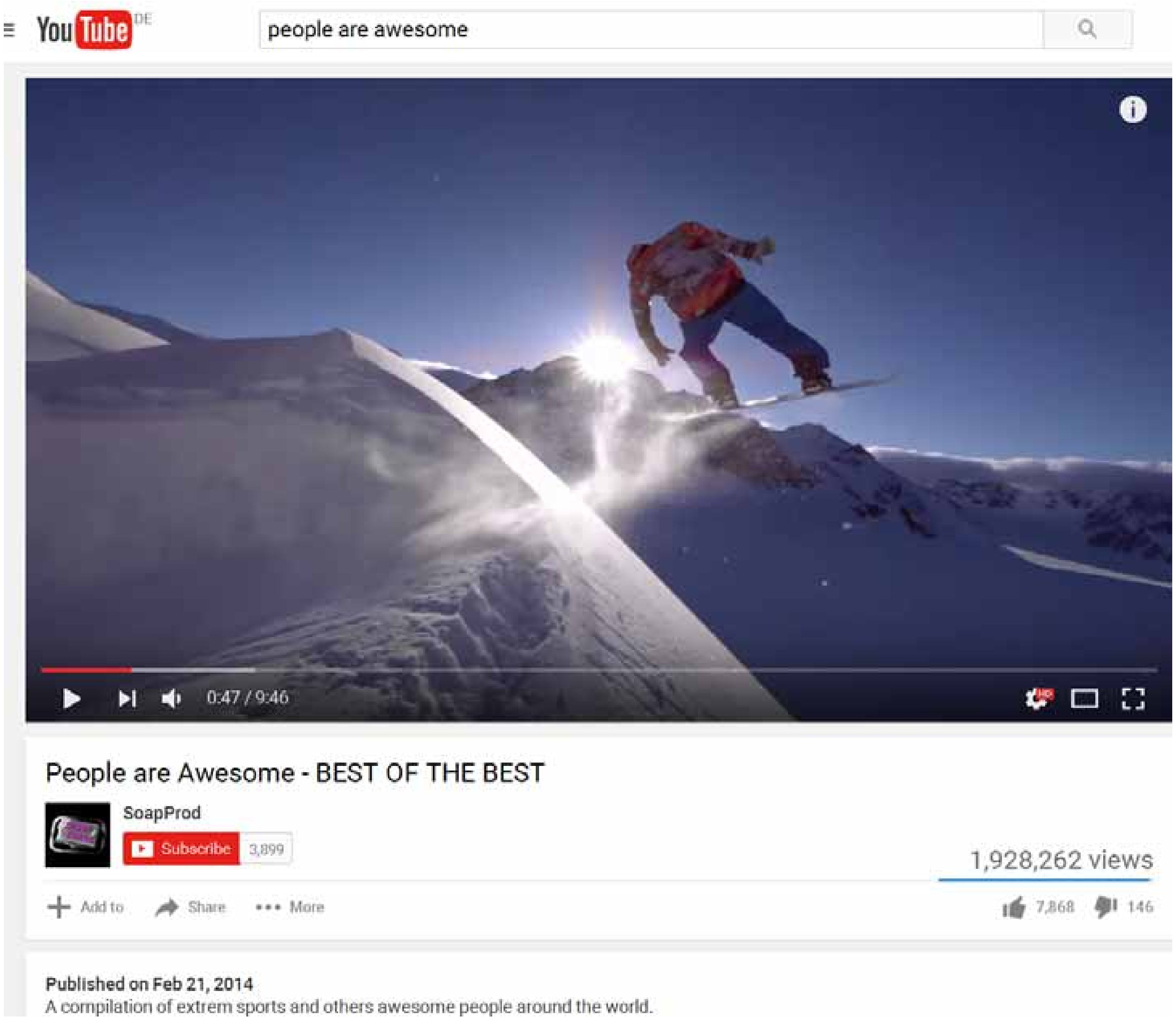
Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.1192788.

Griffiths, T. L. Connecting human and machine learning via probabilistic models of cognition. Interspeech 2009, 2009 Brighton (UK). ISCA, 9-12. available online via: https://cocosci.berkeley.edu/tom/papers/probmods.pdf


Economics
Cognitive Science
Reinforcement Learning
Computer Science
Mathematics

02 When is the human better than a computer?

People are awesome …



See Youtube: "people are awesome" … hundreds of examples

Problem Solving: Humans vs. Computers

When is the human *) better?

*) human intelligence/natural intelligence/human mind/human brain/human learning

- **Natural Language Translation/Curation**
  Machine cannot understand the context of sentences [3]
- **Unstructured problem solving**
  Without a pre-set of rules, a machine has trouble solving the problem, because it lacks the creativity required for it [1]
- **NP-hard Problems**
  Processing times are exponential and makes it almost impossible to use machines for it, so human still stays better [4]

When is the computer **) better?

**) Computational intelligence, Artificial Intelligence/soft computing Machine Learning algorithms

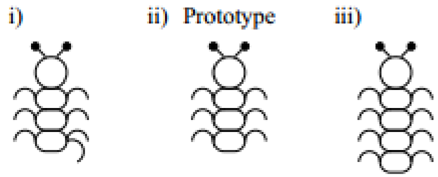
- **High-dimensional data processing**
  Humans are very good at dimensions less or equal than 3, but computers can process data in arbitrarily high dimensions
- **Rule-Based environments**
  Difficulties for humans in rule-based environments often come from not recognizing the correct goal in order to select the correct procedure or set of rules [2]
- **Image optimization**
  Machine can look at each pixel and apply changes without human personal biases, and with more speed [1]

[1] https://www.instartlogic.com/blog/man-vs-machine-learning-based-optimizations

[2] Cummings, Mary Missy. "Man versus machine or man+ machine?." *Intelligent Systems, IEEE* 29.5 (2014): 62-69.

[3] Pizlo, Zygmunt, Anupam Joshi, and Scott M. Graham. "Problem Solving in Human Beings and Computers (formerly: Heuristic Problem Solving)." (1994).

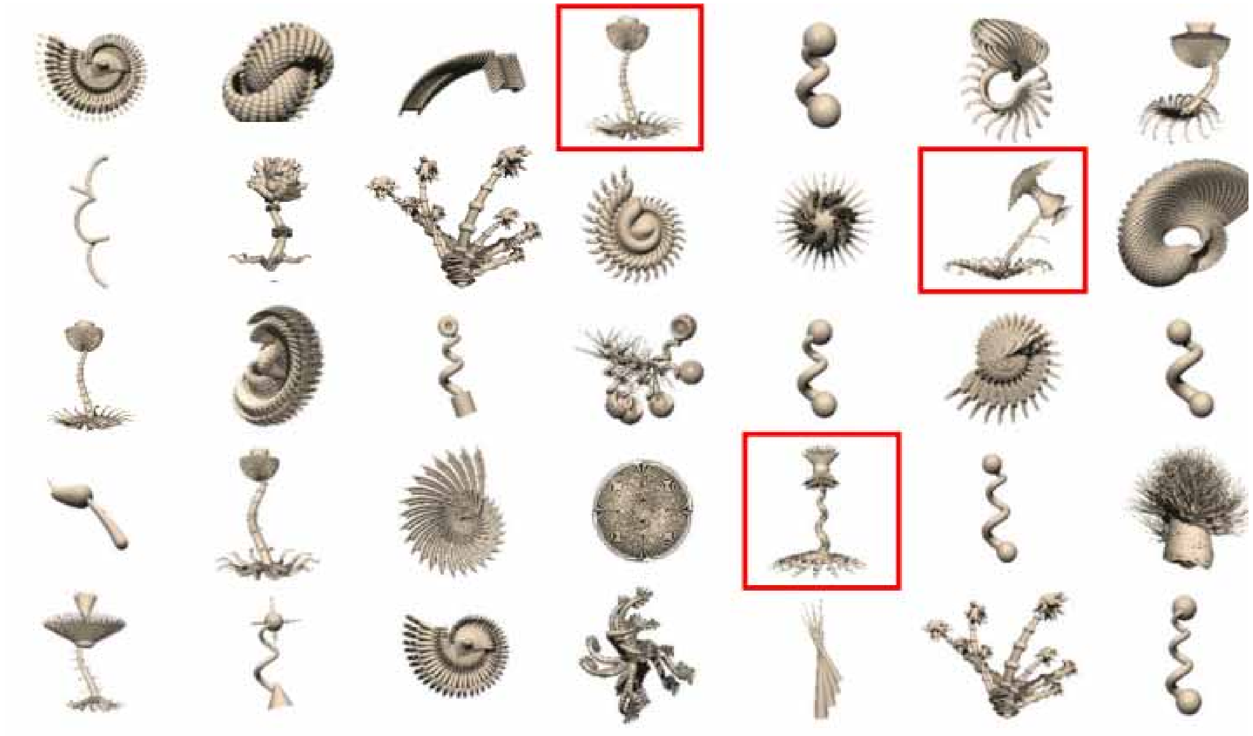
[4] Griffiths, Thomas L. "Connecting human and machine learning via probabilistic models of cognition." *INTERSPEECH*. 2009.



- Similarity [1]
- Representativeness and evidential support
- Causal judgment [2]
- Coincidences and causal discovery
- Clinical diagnostic inference [3]
- Predicting the future

[1] Kemp, C., Bernstein, A. & Tenenbaum, J. B. A generative theory of similarity. Proceedings of the 27th Annual Conference of the Cognitive Science Society, 2005. 1132-1137.
[2] Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J. & Blum, B. 2003. Inferring causal networks from observations and interventions. Cognitive science, 27, (3), 453-489.
[3] Krynski, T. R. & Tenenbaum, J. B. 2007. The role of causality in judgment under uncertainty. Journal of Experimental Psychology: General, 136, (3), 430.

A central question in cognitive science



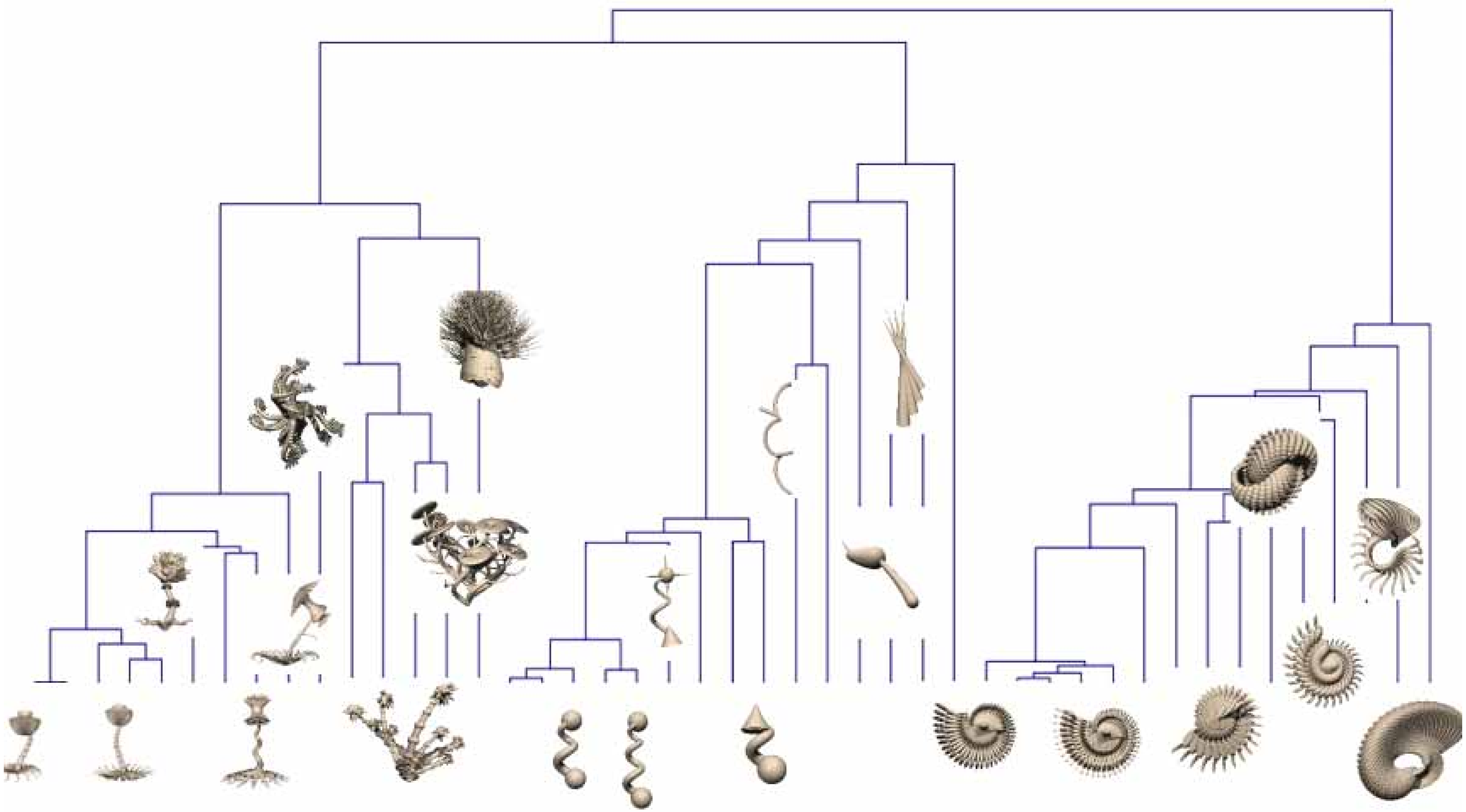
- How does the human mind get so much out of so little?
- Our minds build rich models of the world and make strong generalizations from input data that is sparse, noisy, and ambiguous – in many ways far too limited to support the inferences we make.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.1192788.

Humans learn from very few examples …

Xu, F. & Tenenbaum, J. B. 2007. Word learning as Bayesian inference. Psychological review, 114, (2), 245-272, doi:10.1037/0033-295X.114.2.245.

Humans learn from very few examples …



Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.1192788.

Important: Statistics meet Knowledge

- 1. How does abstract knowledge guide learning and inference from sparse data?
  - (Approximate) Bayesian inference in probabilistic models.
- 2. What are the forms and contents of that knowledge?
  - Probabilities defined over a range of structured representations: graphs, grammars, predicate logic, schemas… programs.
- 3. How is that knowledge itself acquired?
  - Hierarchical Bayesian models, with inference at multiple levels of abstraction ("learning to learn"). Learning as (hierarchical Bayesian) program induction.
- **Central Question:**
  How does our mind get so much out of so little?

Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.1192788.

Example: Discovery of causal relationships from data …

Hans Holbein d.J., 1533,
The Ambassadors,
London: National Gallery

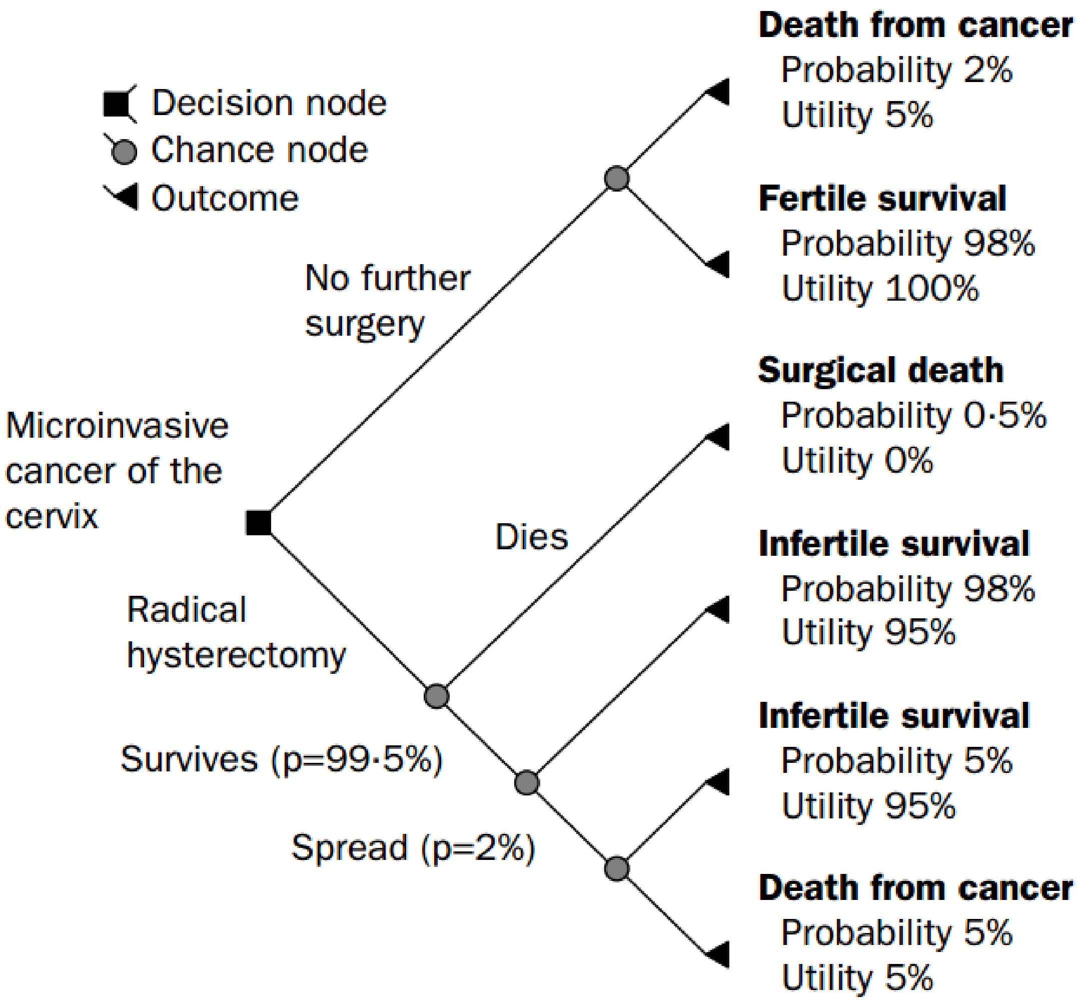
Lopez-Paz, D., Muandet, K., Schölkopf, B. & Tolstikhin, I. 2015. Towards a learning theory of cause-effect inference. Proceedings of the 32nd International Conference on Machine Learning, JMLR, Lille, France.

https://www.youtube.com/watch?v=9KiVNIUMmCc

Is probability an appropriate basis for model cognition?

- Previously this was denied, e.g.: Kahneman & Tversky “Heuristics and biases” 2002 Nobel Prize in Economics: “People are not Bayesian.”
- Slovic, Fischhoff & Lichtenstein (1976): “It appears that people lack the correct programs for many important judgmental tasks ... it may be argued that we have not had the opportunity to evolve an intellect capable of dealing conceptually with uncertainty.”
- Stephen J.Gould (1992): “Our minds are not built (for whatever reason) to work by the rules of probability” …

Decision trees are coming from Clinical Practice



Physician treating a patient approx. 480 B.C.
Beazley (1963), Attic Red-figured Vase-Painters, 813, 96.
Department of Greek, Etruscan and Roman Antiquities, Sully, 1st floor, Campana Gallery, room 43 Louvre, Paris

Elwyn, G., Edwards, A., Eccles, M. & Rovner, D. 2001. Decision analysis in patient care. The Lancet, 358, (9281), 571-574.

For a single decision variable an agent can select $D = d$ for any $d \in dom(D)$.

The expected utility of decision $D = d$ is

http://www.eoht.info/page/Oskar+Morgenstern

$$E(U \mid d) = \sum_{x_1,\ldots,x_n} P(x_1,\ldots,x_n \mid d) U(x_1,\ldots,x_n,d)$$

An optimal single decision is the decision $D = dmax$ whose expected utility is maximal:

$$d_{\max} = \arg \max_{d \in \mathrm{dom}(D)} E(U \mid d)$$

Von Neumann, J. & Morgenstern, O. 1947. Theory of games and economic behavior, Princeton university press.

Key Challenges

- Medicine is an extremely complex application domain – dealing most of the time with uncertainties -> **probable information!**
- Key: Structure learning and prediction in large-scale biomedical networks with probabilistic graphical models
- Causal and Probabilistic Inference:
- Uncertainties are present at all levels in health related systems
- Data sets from which ML learns are noisy, mislabeled, atypical, etc. etc.
- Even with data of high quality, gauging and combining a multitude of data sources and constraints in usually imperfect models of the world requires us to represent and process uncertain knowledge in order to make viable decisions.
- In the increasingly complicated settings of modern science, model structure or causal relationships may not be known a-priori [1].
- Approximating probabilistic inference in Bayesian belief networks is NP-hard [2] -> here we need the "human-in-the-loop" [3]

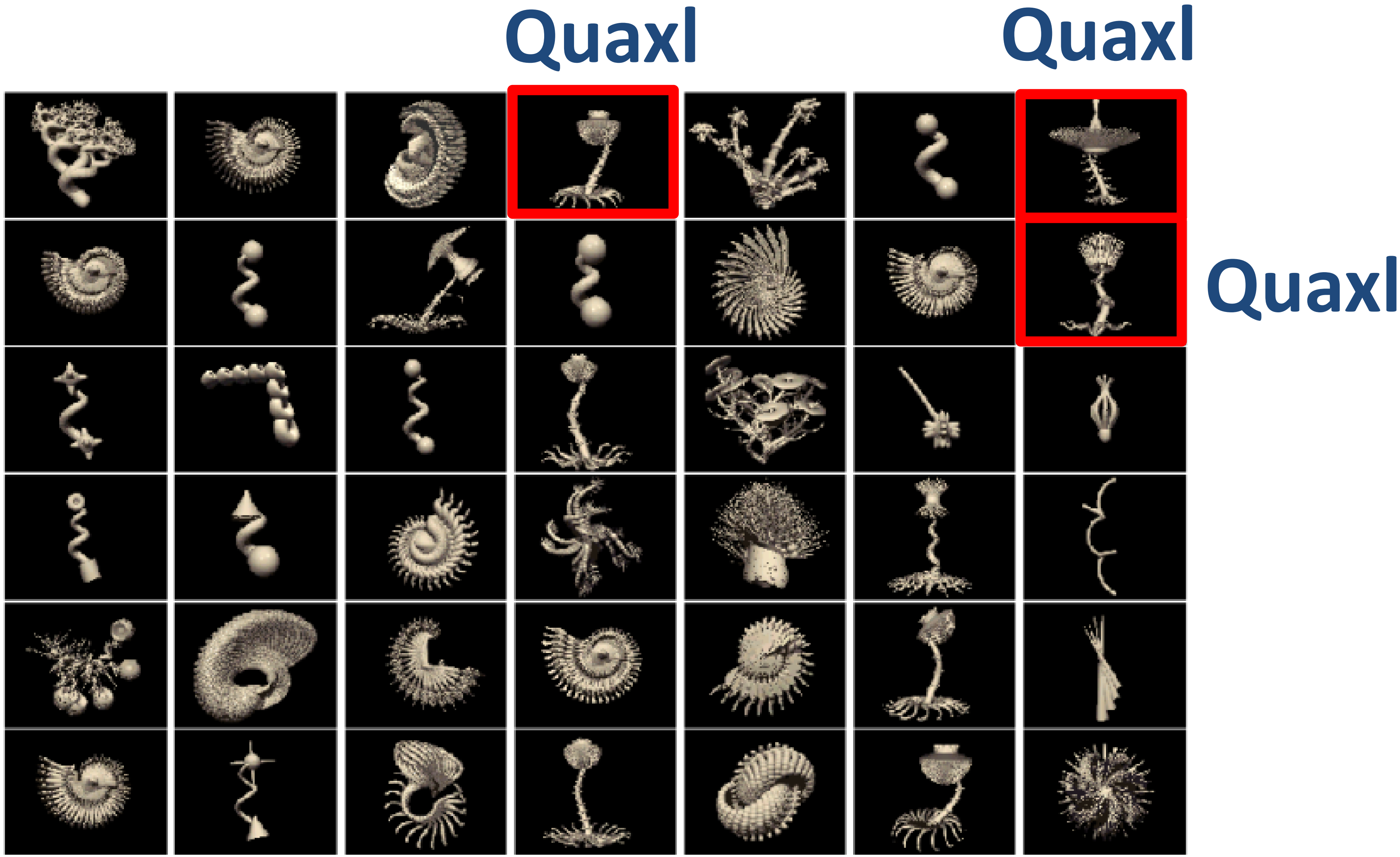
[1] Sun, X., Janzing, D. & Schölkopf, B. Causal Inference by Choosing Graphs with Most Plausible Markov Kernels. ISAIM, 2006.
[2] Dagum, P. & Luby, M. 1993. Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artificial intelligence, 60, (1), 141-153.
[3] Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Springer Brain Informatics (BRIN), 3, 1-13, doi:10.1007/s40708-016-0042-6.

03 Human Information Processing

How does our mind get so much out of it …

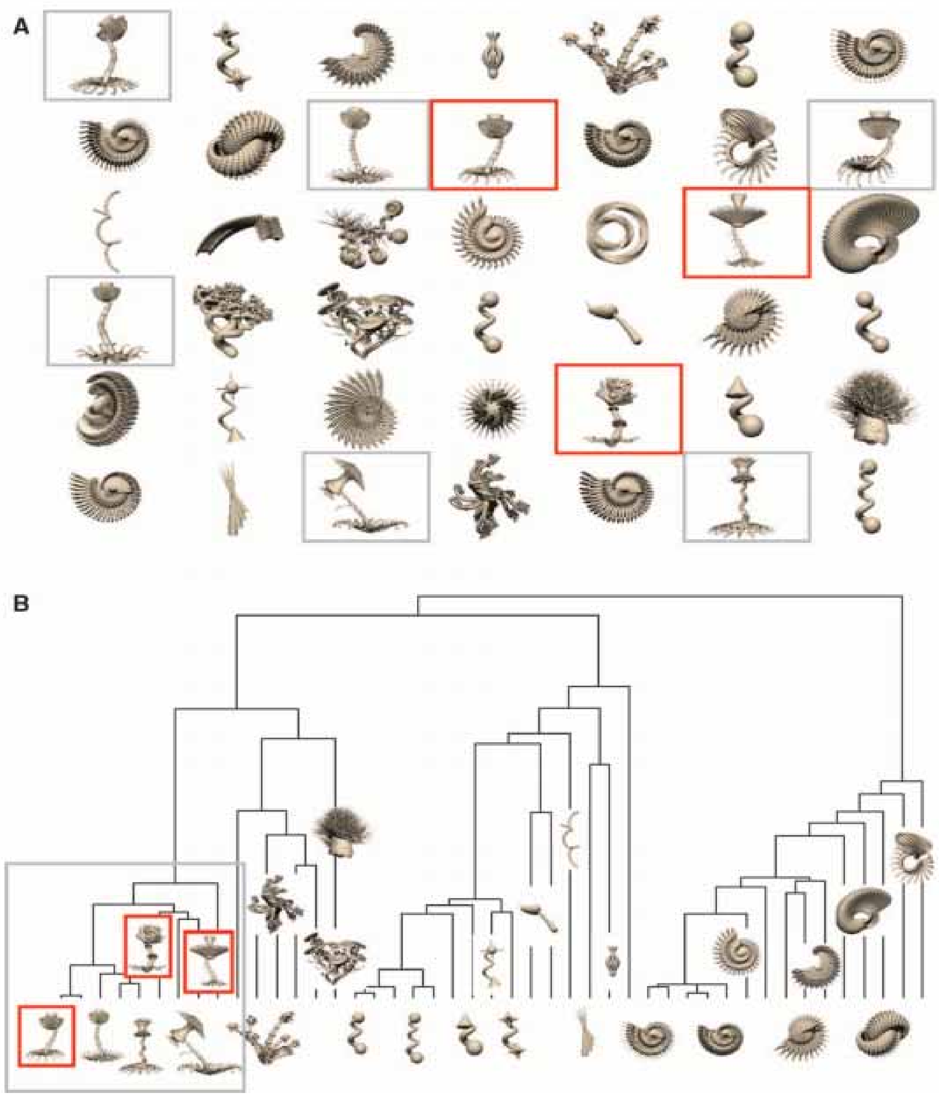
Salakhutdinov, R., Tenenbaum, J. & Torralba, A. 2012. One-shot learning with a hierarchical nonparametric Bayesian model. Journal of Machine Learning Research, 27, 195-207.

Learning words for objects – concepts from examples

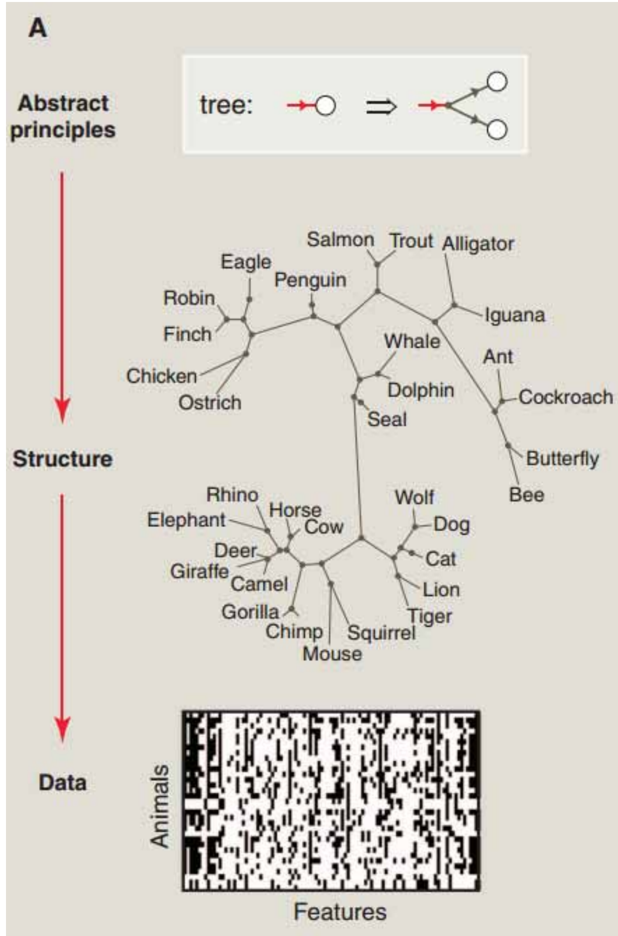


Salakhutdinov, R., Tenenbaum, J. & Torralba, A. 2012. One-shot learning with a hierarchical nonparametric Bayesian model. Journal of Machine Learning Research, 27, 195-207.

How do we understand our world …



$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h' \in H} P(d|h')P(h')} \propto P(d|h)P(h)$$

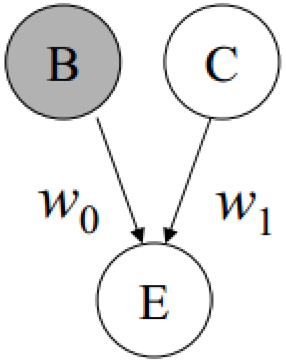


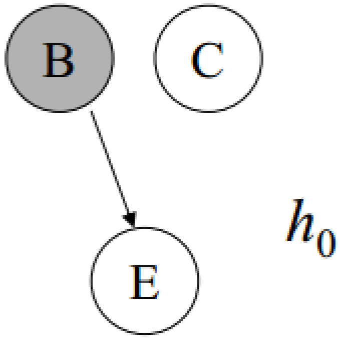
Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285.

One of the unsolved problems in human concept learning

- which is highly relevant for ML research, concerns the factors that determine the subjective difficulty of concepts:
- Why are some concepts psychologically extremely simple and easy to learn,
- while others seem to be extremely difficult, complex, or even incoherent?
- These questions have been studied since the 1960s but are still unanswered …

Feldman, J. 2000. Minimization of Boolean complexity in human concept learning. Nature, 407, (6804), 630-633, doi:10.1038/35036586.



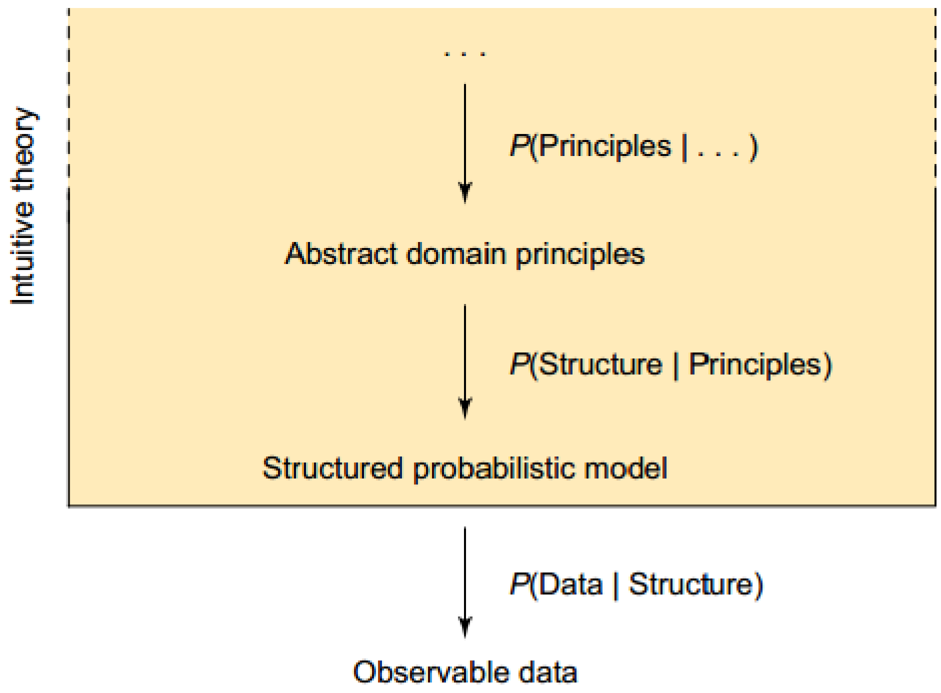


- Cognition as probabilistic inference
  - Visual perception, language acquisition, motor learning, associative learning, memory, attention, categorization, reasoning, causal inference, decision making, theory of mind
- Learning concepts from examples
- Learning and applying intuitive theories (balancing complexity vs. fit)

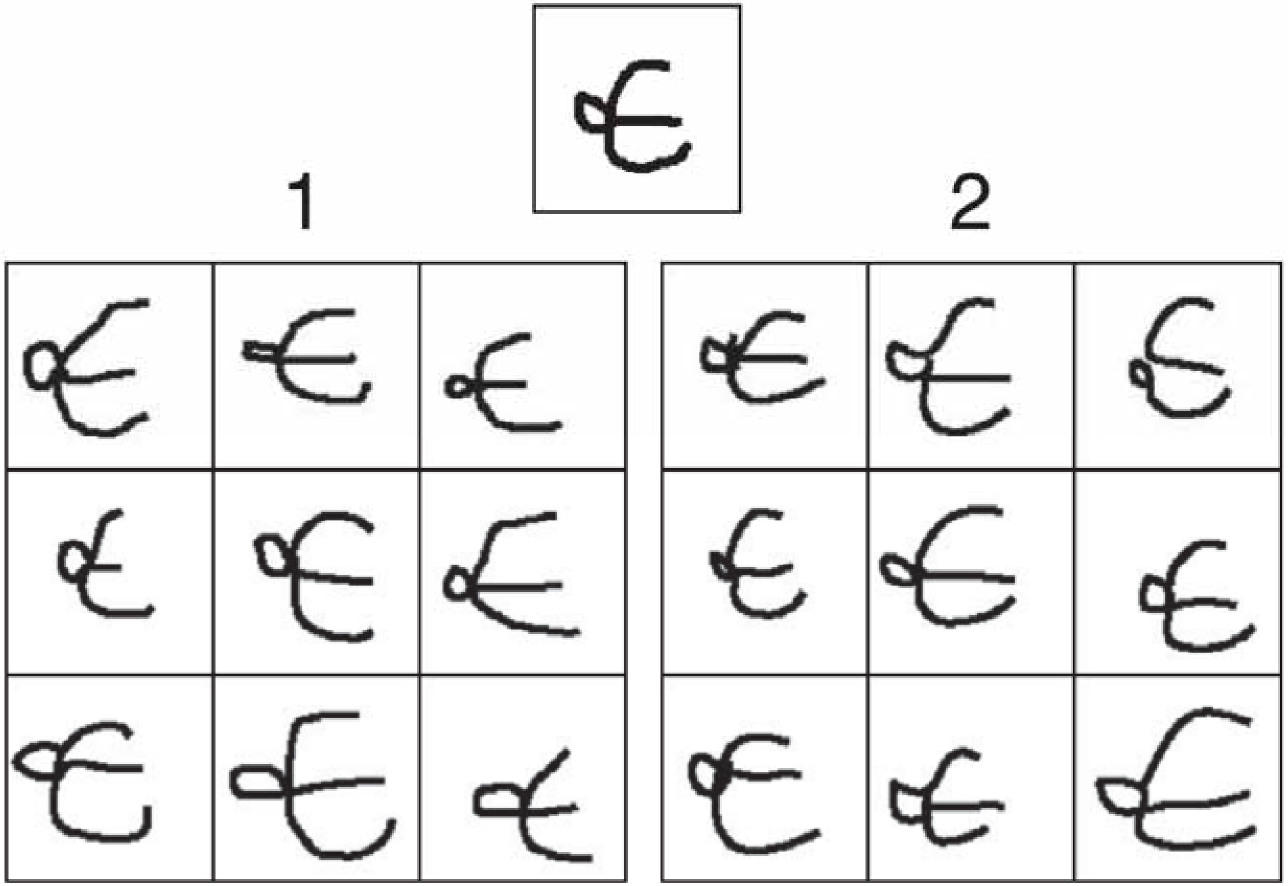
Modeling basic cognitive capacities as intuitive Bayes

- Similarity
- Representativeness and evidential support
- Causal judgement
- Coincidences and causal discovery
- Diagnostic inference
- Predicting the future

Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. 2006. Theory-based Bayesian models of inductive learning and reasoning. Trends in cognitive sciences, 10, (7), 309-318.

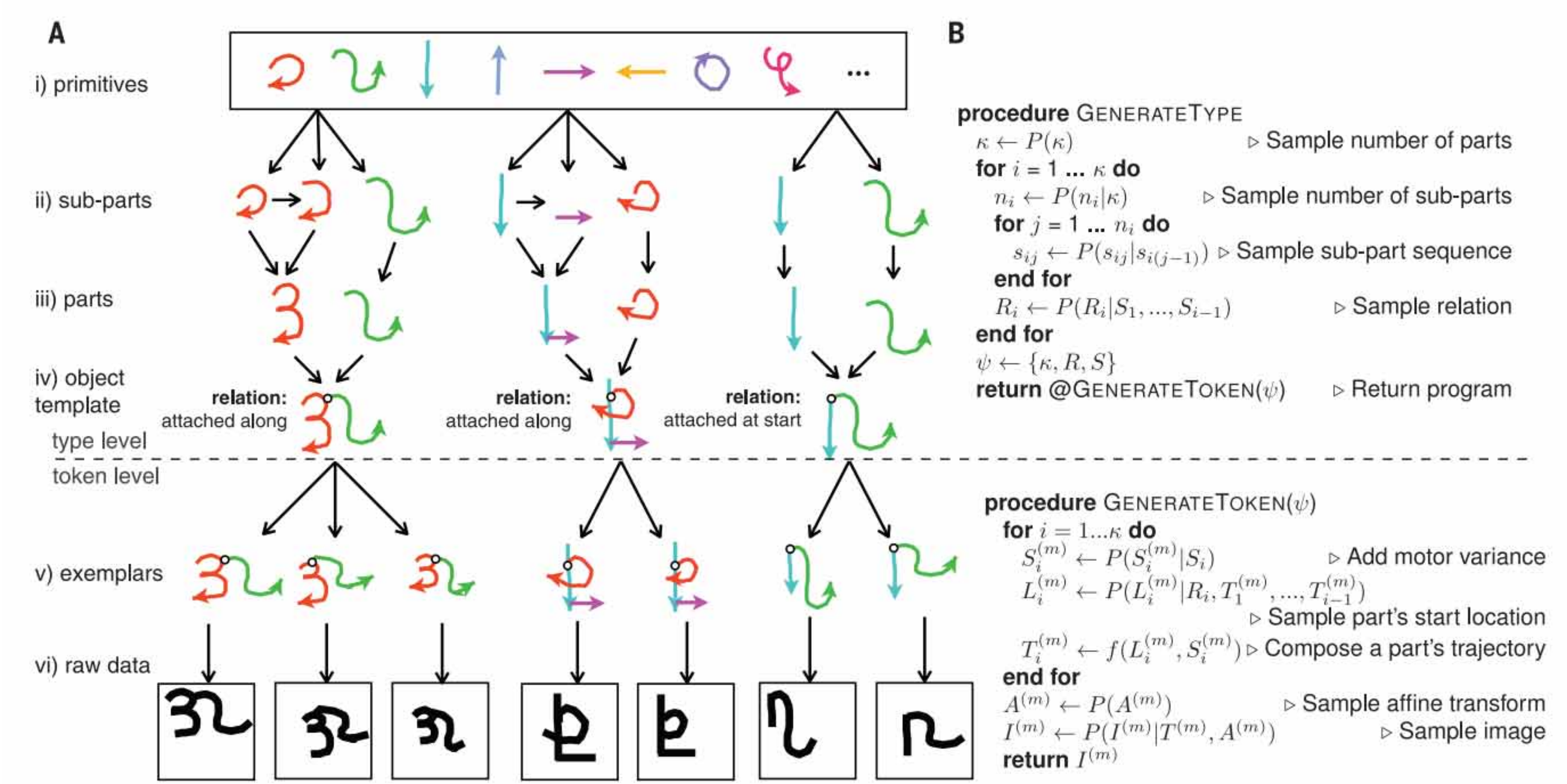


Drawn by Human or Machine Learning Algorithm?



Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. 2015. Human-level concept learning through probabilistic program induction. Science, 350, (6266), 1332-1338, doi:10.1126/science.aab3050.

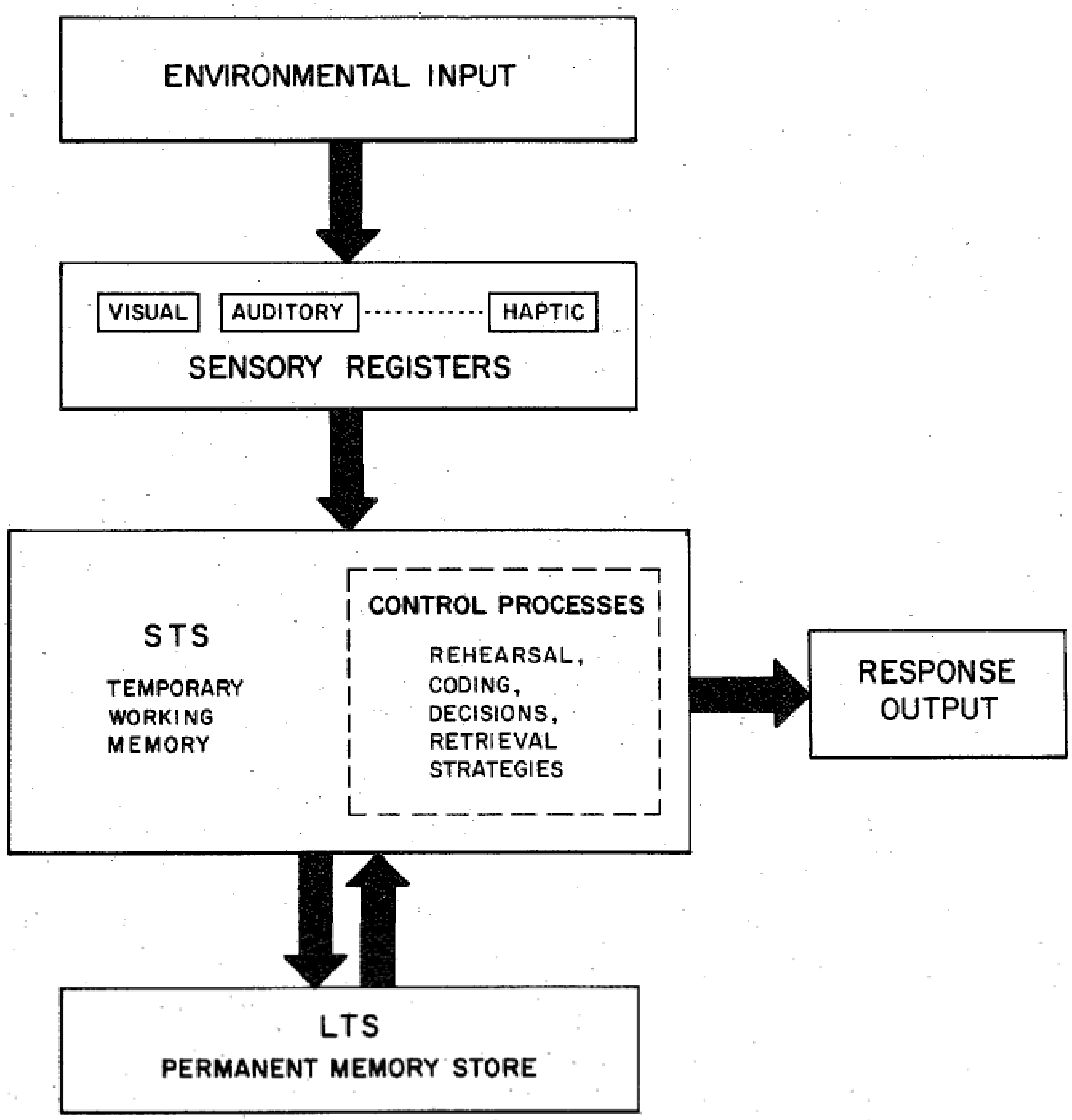
A Bayesian program learning (BPL) framework, capable of learning a large class of visual concepts from just a single example and generalizing in ways that are mostly indistinguishable from people



Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. 2015. Human-level concept learning through probabilistic program induction. Science, 350, (6266), 1332-1338, doi:10.1126/science.aab3050.

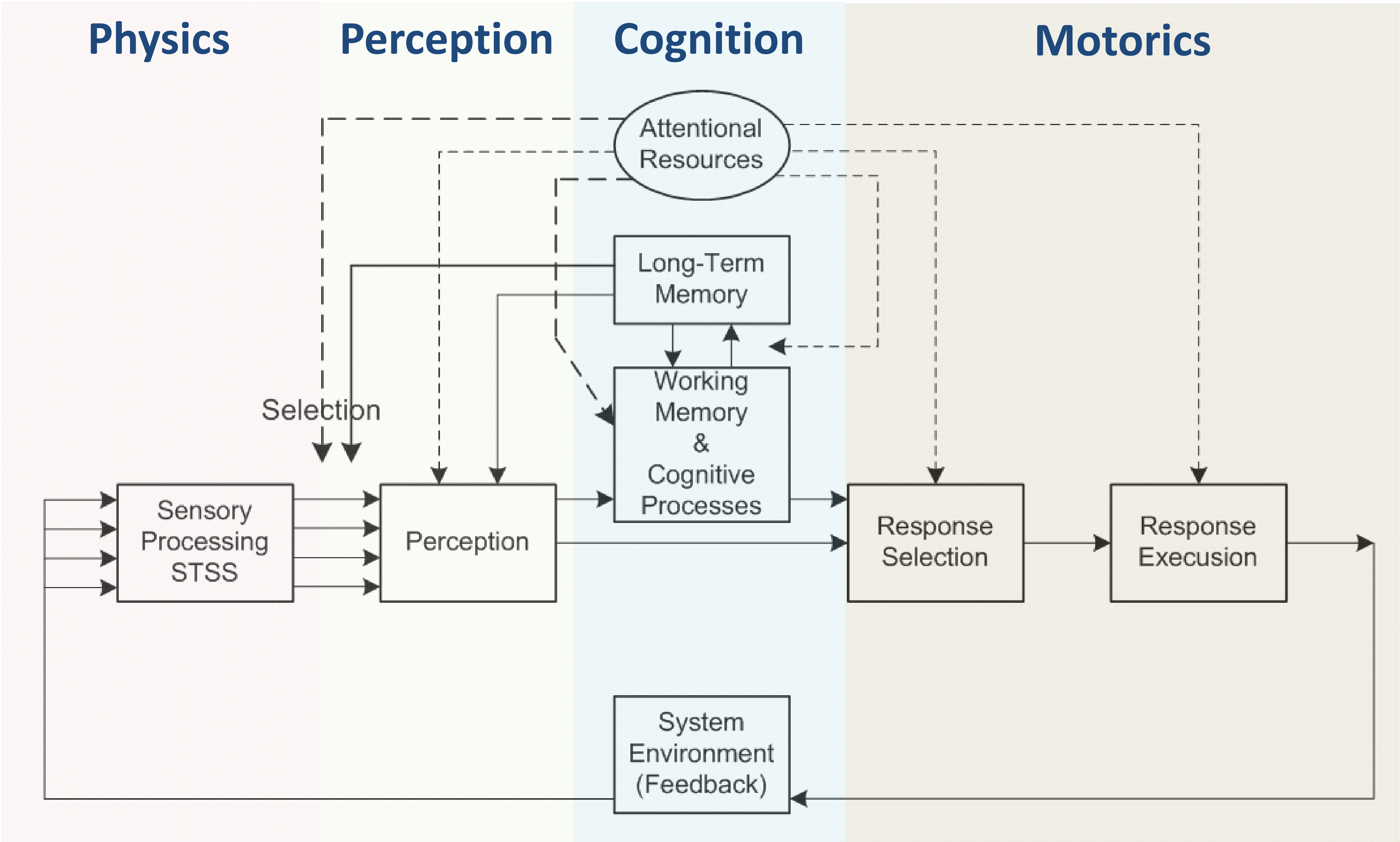
How does our mind get so much out of so little?

Human Information Processing Model (A&S)



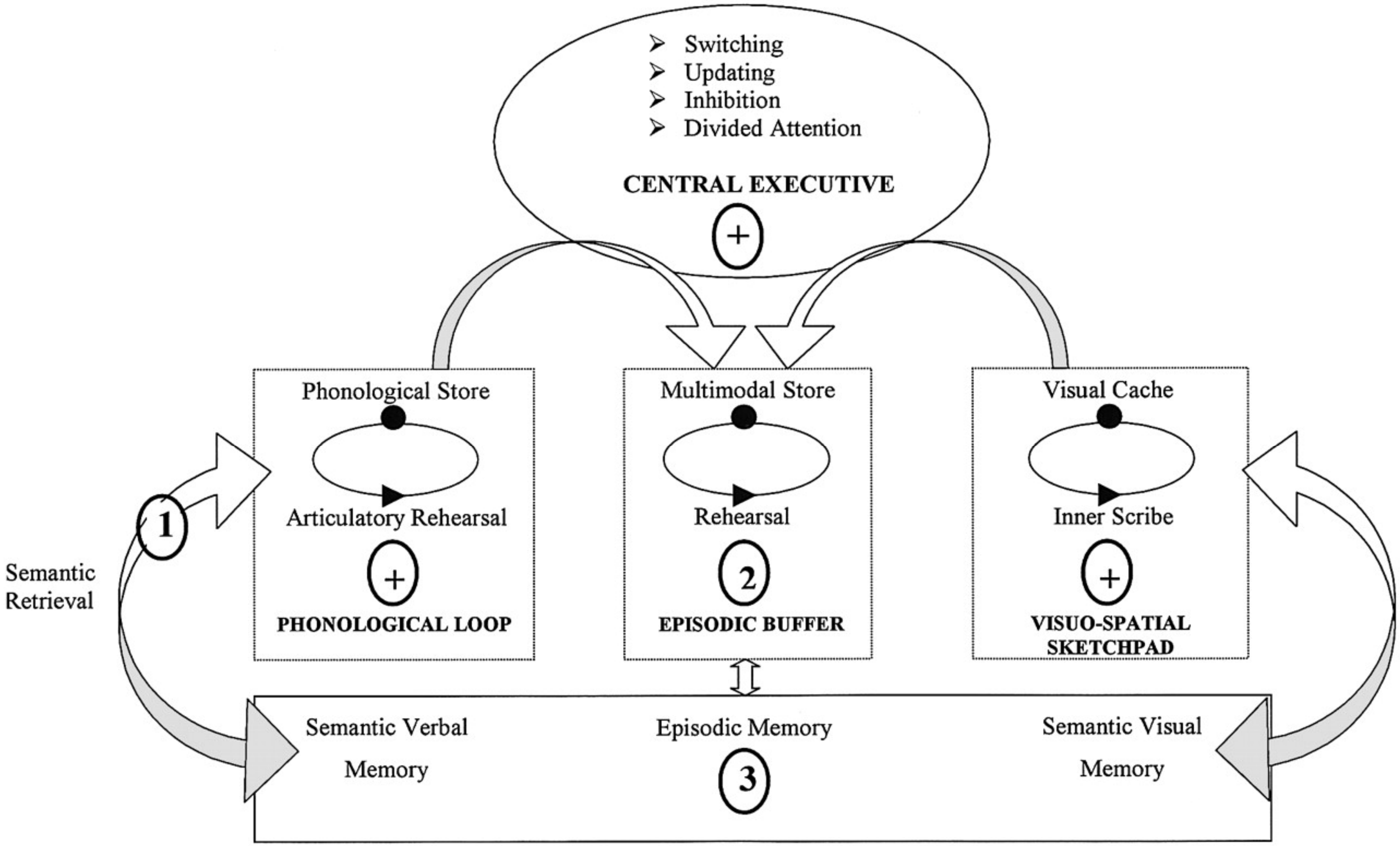
Atkinson, R. C. & Shiffrin, R. M. (1971) *The control processes of short-term memory (Technical Report 173, April 19, 1971). Stanford, Institute for Mathematical Studies in the Social Sciences, Stanford University.*

General Model of Human Information Processing



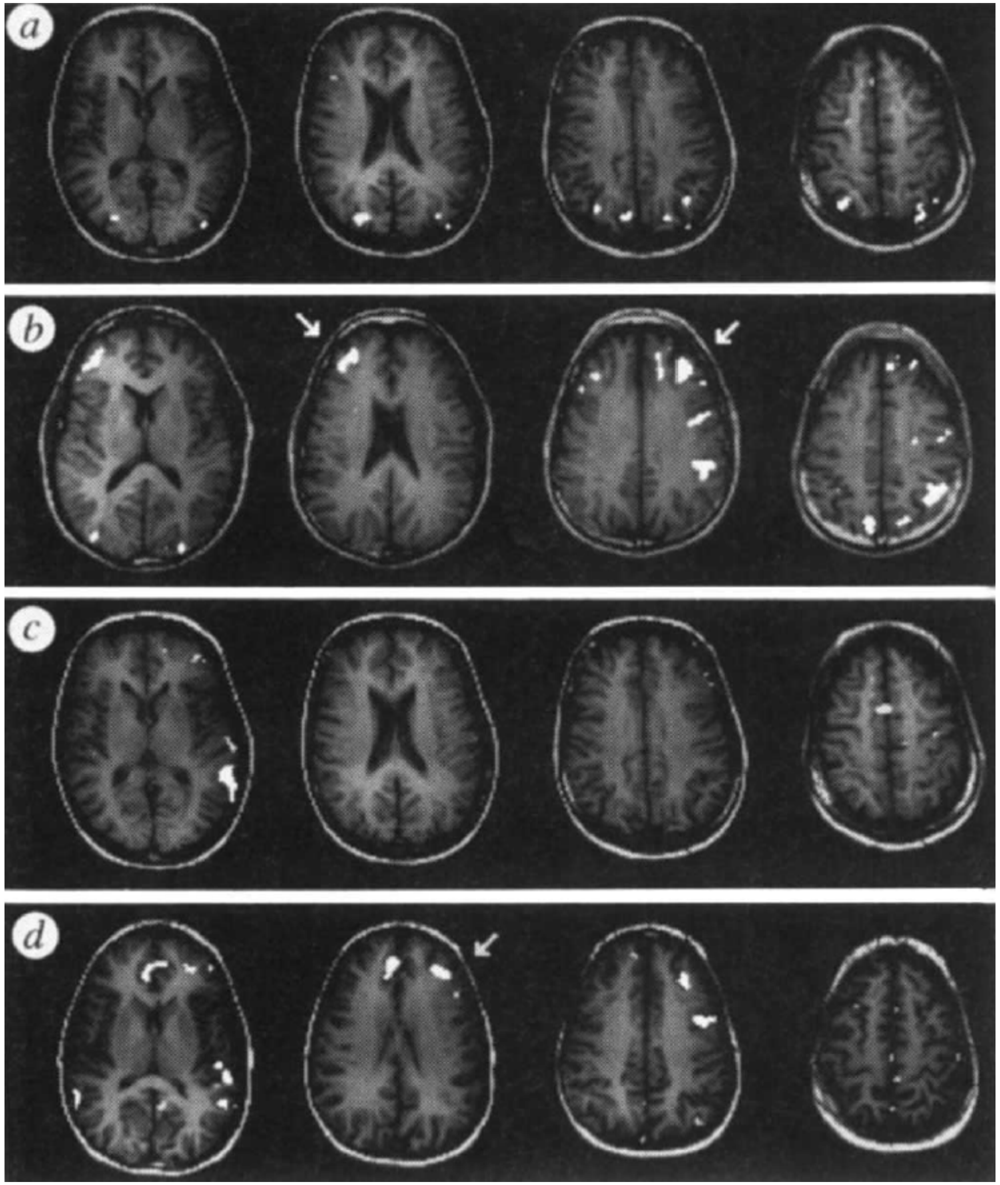
Wickens, C., Lee, J., Liu, Y. & Gordon-Becker, S. (2004) *Introduction to Human Factors Engineering: Second Edition. Upper Saddle River (NJ), Prentice-Hall.*

Alternative Model: Baddeley - Central Executive

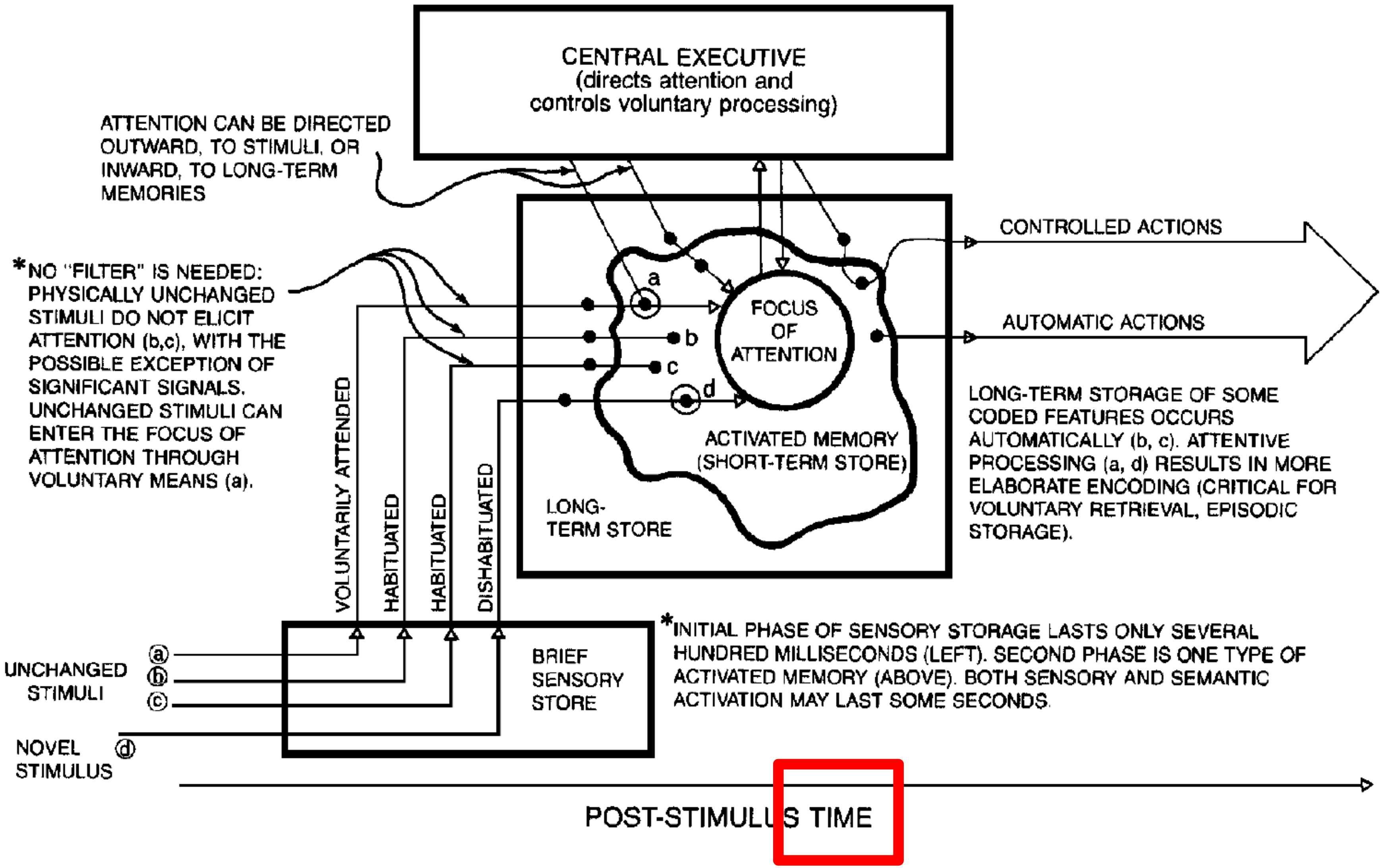


Quinette, P., Guillery, B., Desgranges, B., de la Sayette, V., Viader, F. & Eustache, F. (2003) Working memory and executive functions in transient global amnesia. *Brain, 126, 9, 1917-1934.*

Neural Basis for the "Central Executive System"



D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S. & Grossman, M. (1995) The neural basis of the central executive system of working memory. *Nature, 378, 6554, 279-281.*



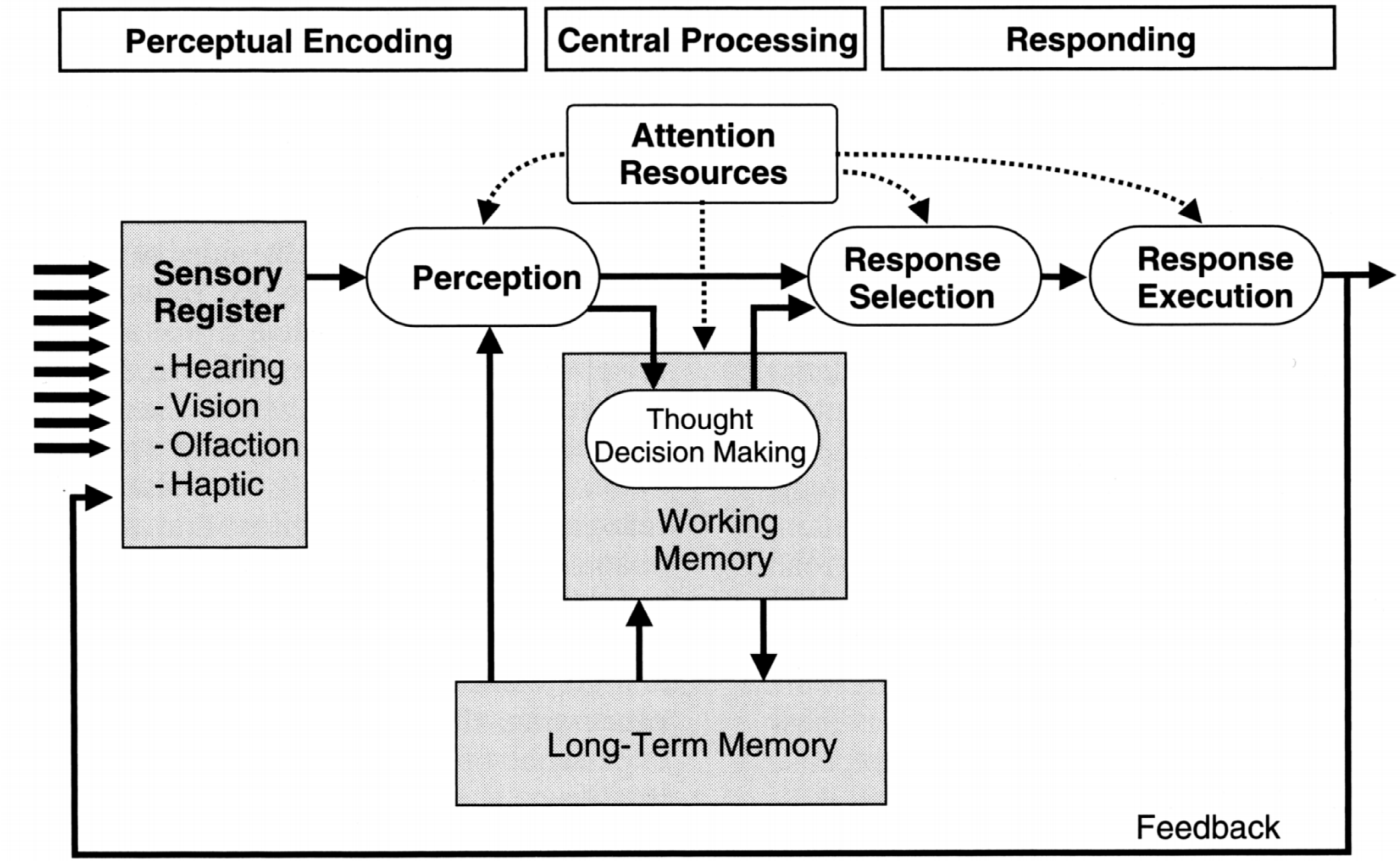
Cowan, N. (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104, 2, 163.*

Selective Attention Test

Note: The Test does NOT properly work if you know it in advance or if you do not concentrate on counting

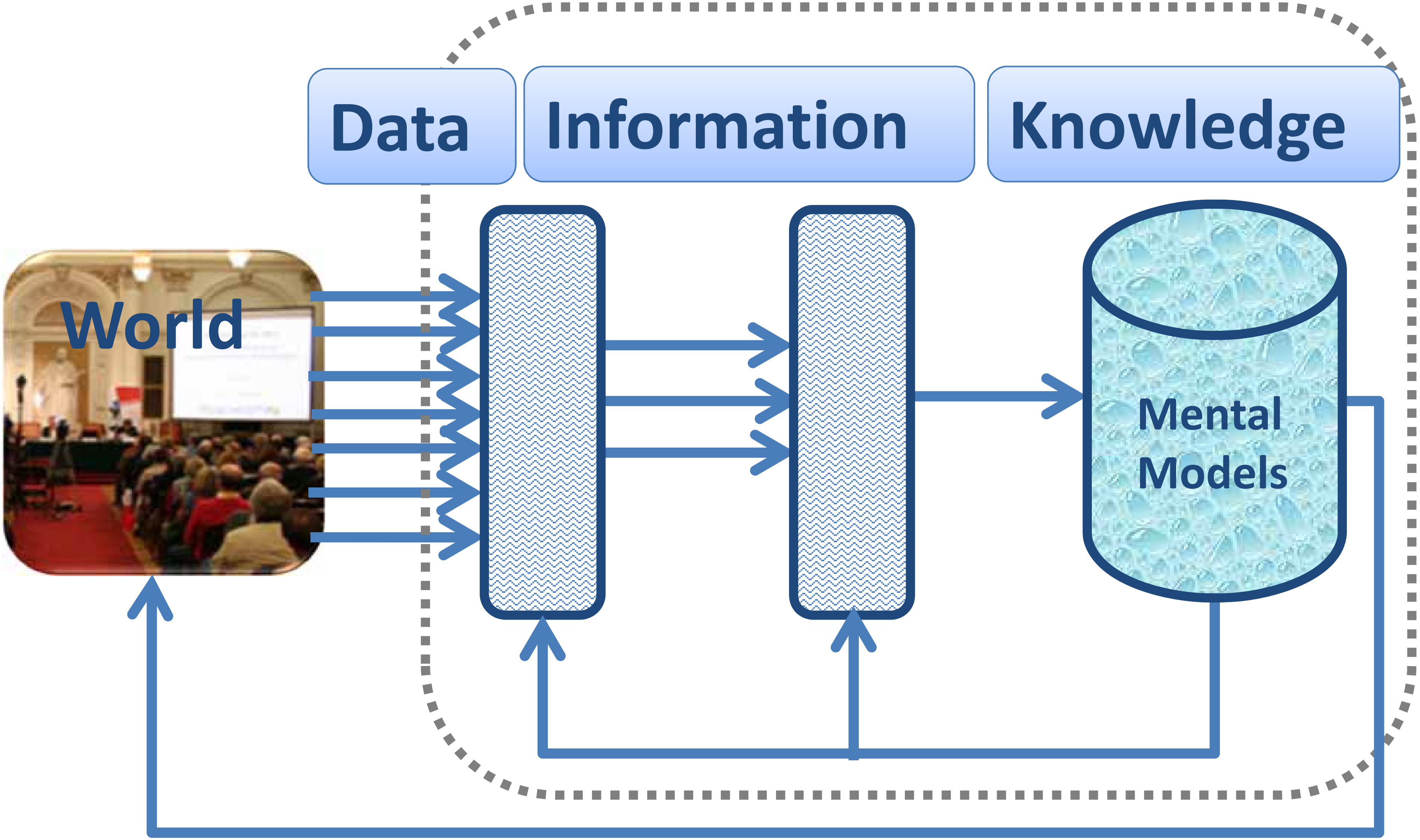
Simons, D. J. & Chabris, C. F. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception, 28, (9), 1059-1074.

Human Attention is central for decision making



Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill.*

Our definition of Knowledge – adaptive agent



Knowledge := a set of expectations

04 Decision Making under Uncertainty

Decision Making is central in Health Informatics

Source: Cisco (2008). Cisco Health Presence Trial at Aberdeen Royal Infirmary in Scotland

3 July 1959, Volume 130, Number 3366 SCIENCE

Reasoning Foundations of Medical Diagnosis

Symbolic logic, probability, and value theory aid our understanding of how physicians reason.

Robert S. Ledley and Lee B. Lusted

The purpose of this article is to analyze the complicated reasoning processes inherent in medical diagnosis. The importance of this problem has received recent emphasis by the increasing interest in the use of electronic computers as an aid to medical diagnostic processes

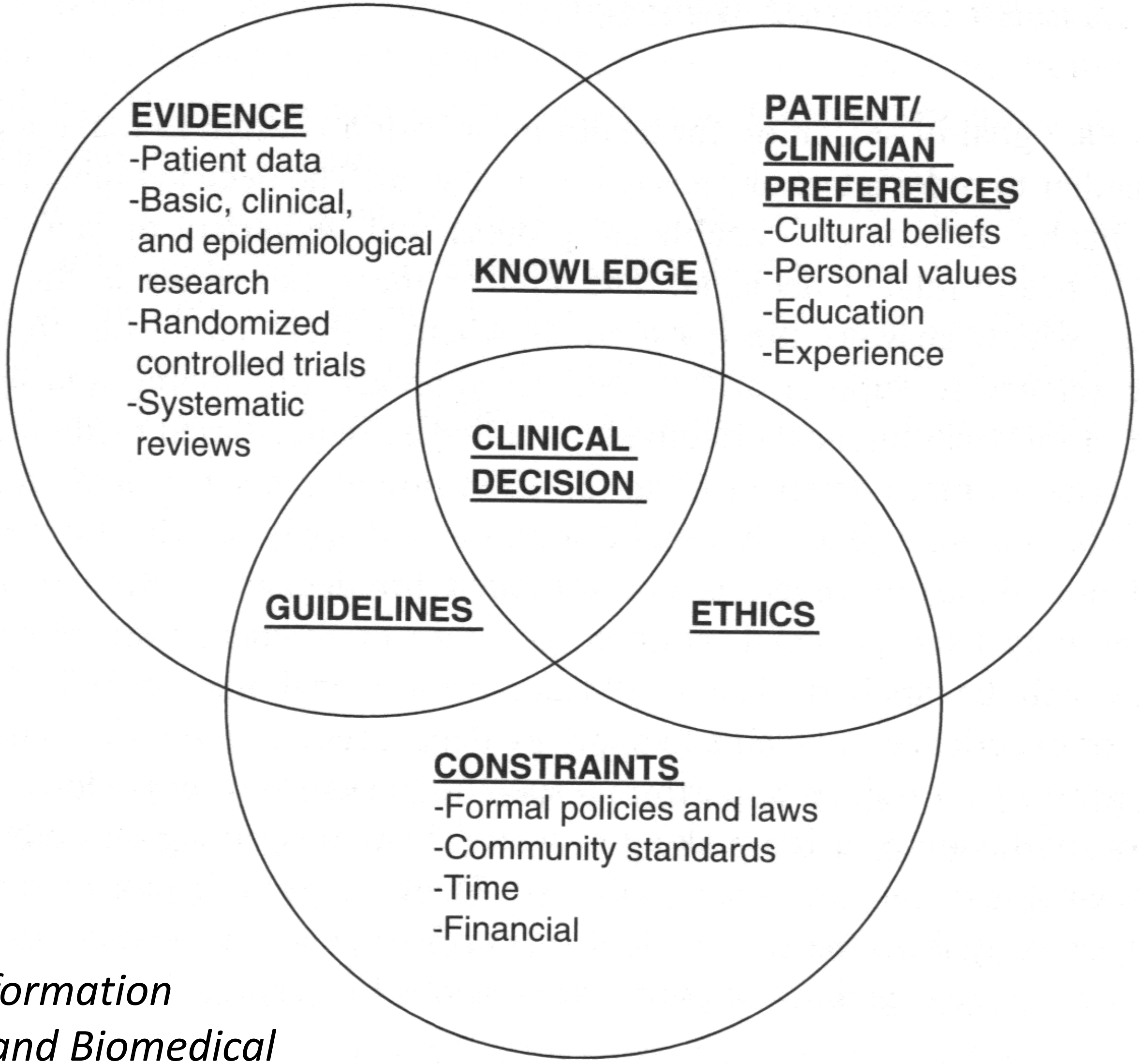
fitted into a definite disease category, or that it may be one of several possible diseases, or else that its exact nature cannot be determined." This, obviously, is a greatly simplified explanation of the process of diagnosis, for the physician might also comment that after seeing a

ance are the ones who do remember and consider the most possibilities."

Computers are especially suited to help the physician collect and process clinical information and remind him of diagnoses which he may have overlooked. In many cases computers may be as simple as a set of hand-sorted cards, whereas in other cases the use of a large-scale digital electronic computer may be indicated. There are other ways in which computers may serve the physician, and some of these are suggested in this paper. For example, medical students might find the computer an important aid in learning the methods of differential diagnosis. But to use the computer thus we must understand how the physician makes a medical diagnosis. This, then, brings us to the subject of our investigation: the reasoning foundations of medical diagnosis and treatment.

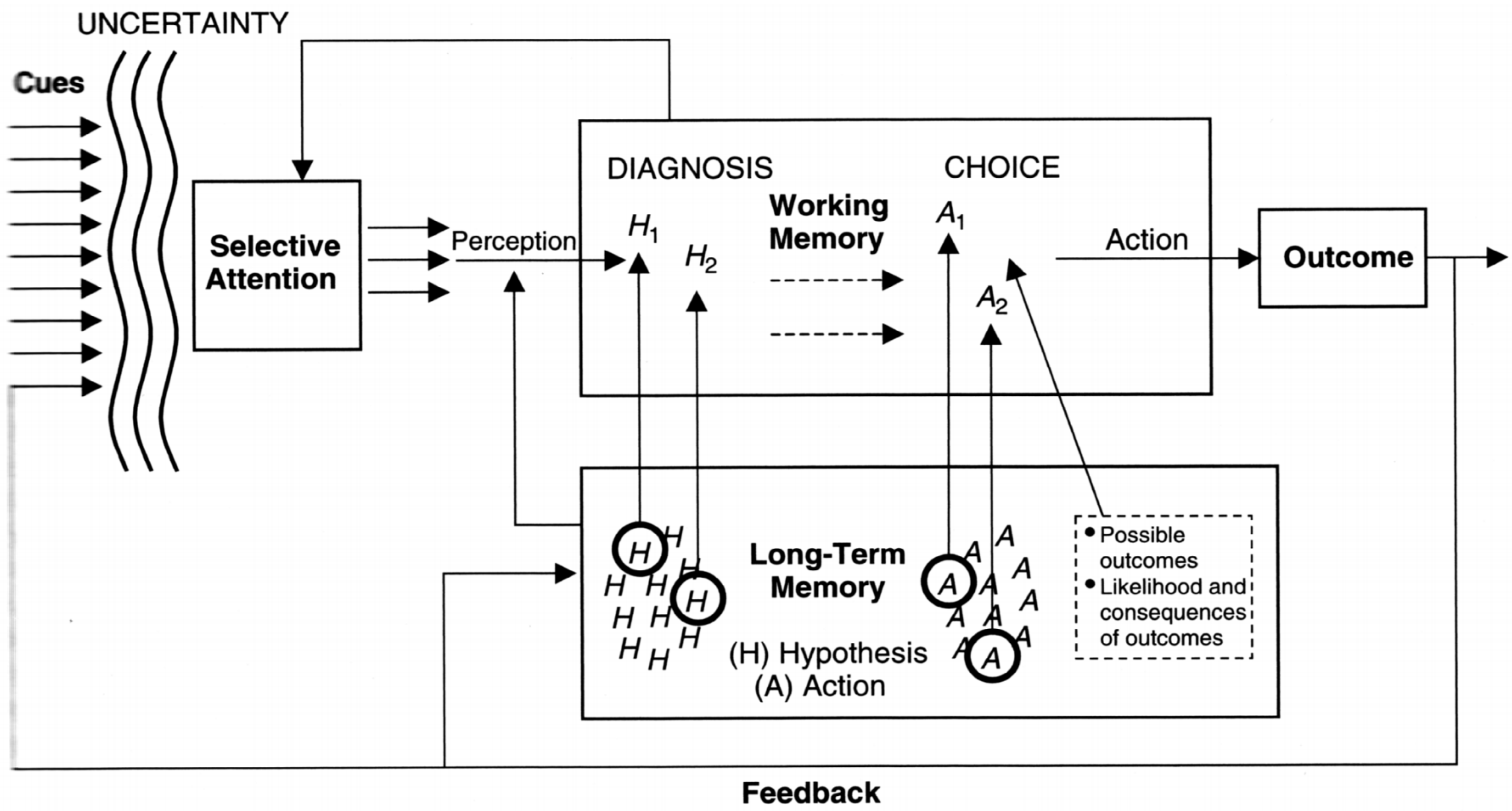
Medical diagnosis involves processes that can be systematically analyzed, as well as those characterized as "intangible." For instance, the reasoning foundations of medical diagnostic procedures

Clinical Medicine is Decision Making!



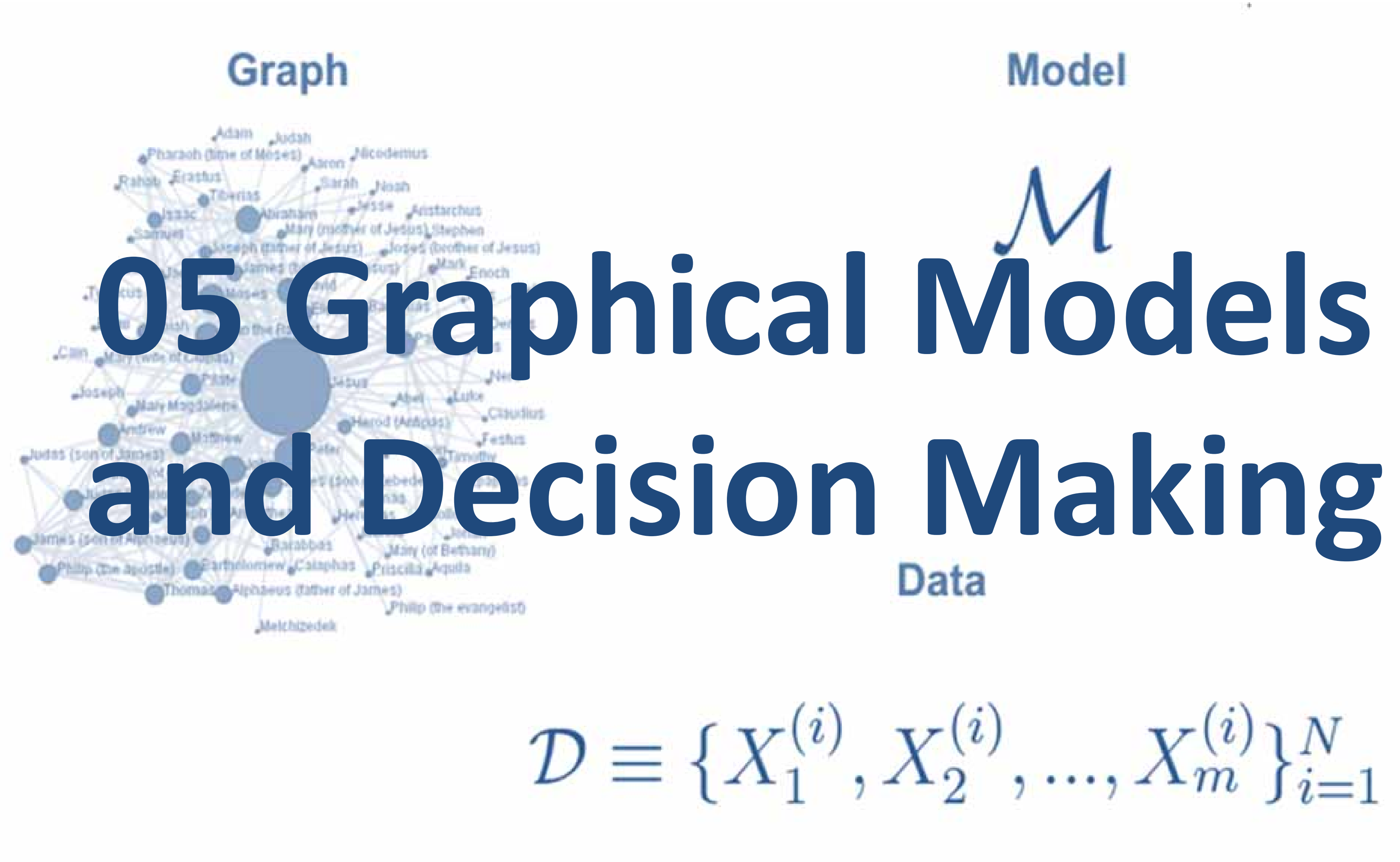
Hersh, W. (2010) *Information Retrieval: A Health and Biomedical Perspective. New York, Springer.*

Human Decision Making



Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill.*


Medical Action ...
is permanent decision making

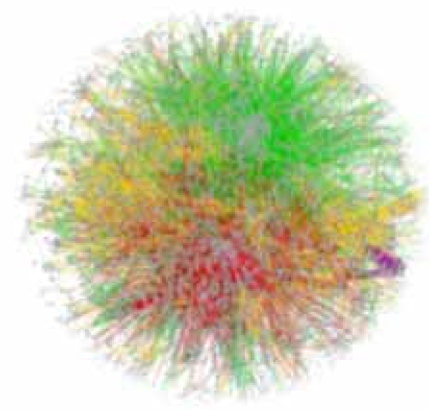


05 Graphical Models and Decision Making

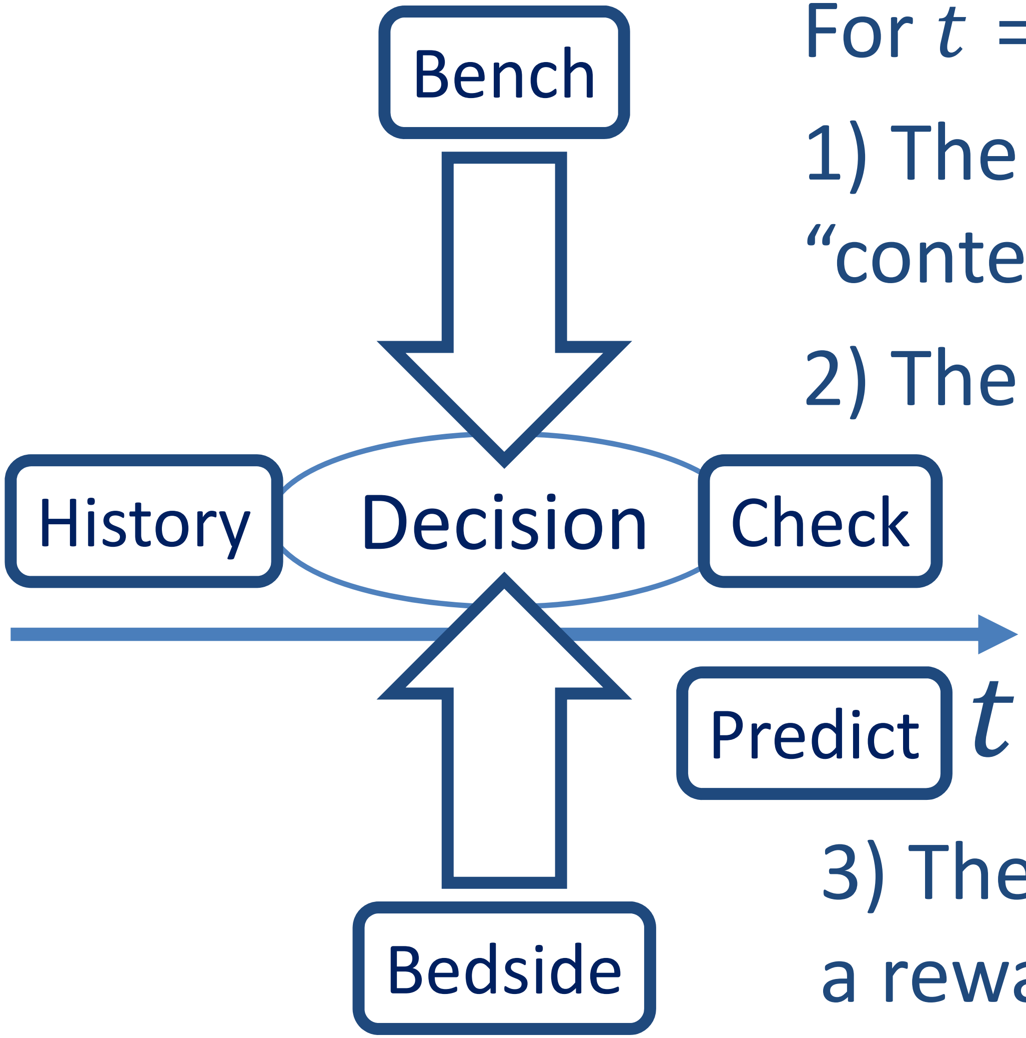
$$\mathcal{D} \equiv \{X_1^{(i)}, X_2^{(i)}, \ldots, X_m^{(i)}\}_{i=1}^{N}$$

What are Probabilistic Graphical Models?

- PGM can be seen as a combination between
- Graph Theory + Probability Theory + Machine Learning
- One of the most exciting advancements in AI in the last decades
- Compact representation for exponentially-large probability distributions
- Example Question:
  "Is there a path connecting two proteins?"
- $Path\ (X,Y) :=\ edge\ (X,Y)$
- $Path\ (X,Y) :=\ edge\ (X,Y), path(Z,Y)$
- This can NOT be expressed in first-order logic
- Need a Turing-complete fully-fledged language



Goal: Learn an **optimal policy** for selecting best actions within a given **context**



For $t = 1, \ldots, T$

1) The world produces a "context" $x_t \in X$

2) The learner selects an action $a_t \in \{1, \ldots, K\}$

3) The world reacts with a reward $r_t(a_t) \in [0,1]$

GM are amongst the most important ML developments

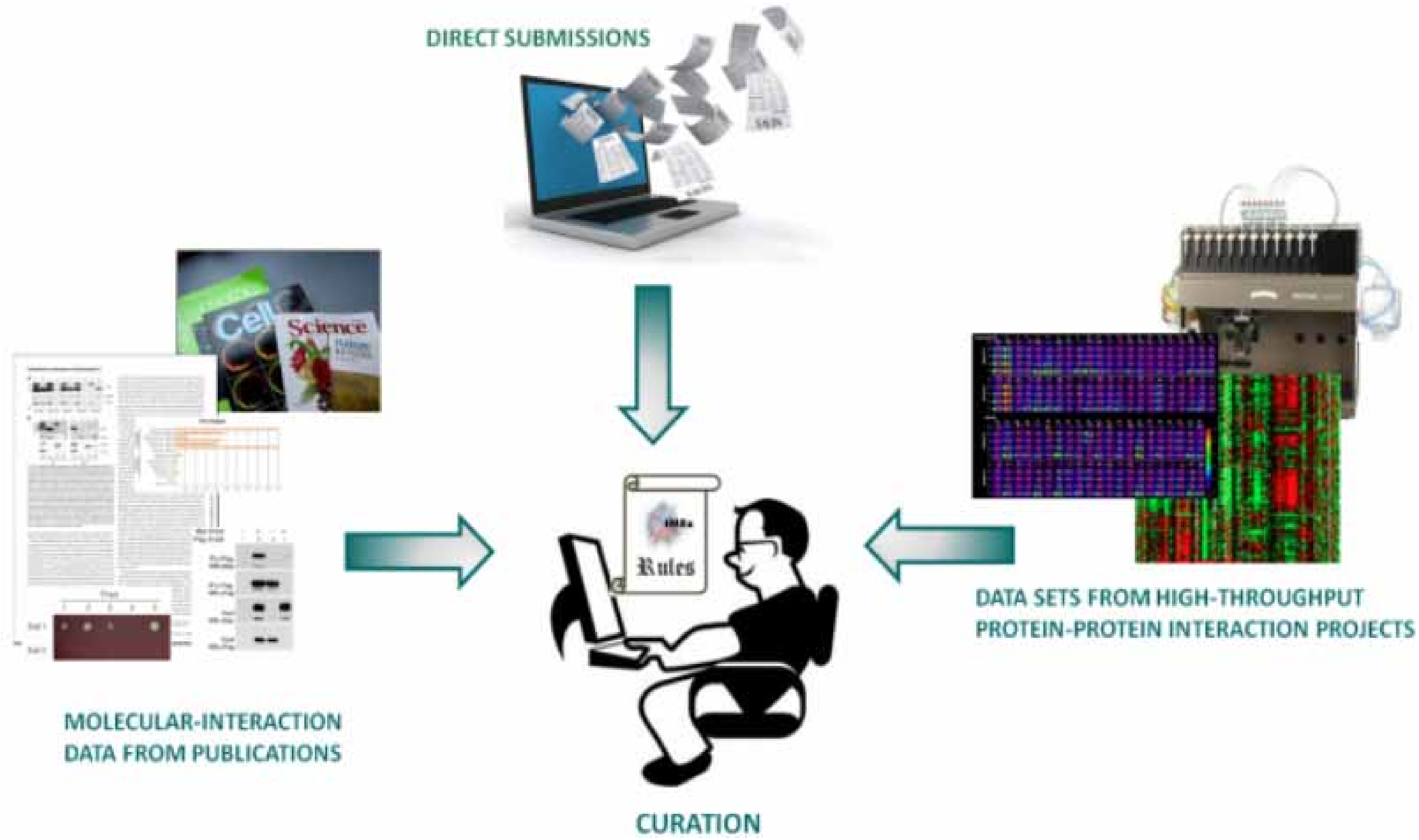
- Key Idea: Conditional independence assumptions are very useful – however: Naïve Bayes is extreme!
- X is *conditionally independent* of Y, given Z, if the P(X) governing X is independent of value Y, given value of Z:

$$(\forall i, j, k) P(X = x_i | Y = y_j, Z = z_k) = P(X = x_i | Z = z_k)$$

can be abbr. with $P(X|Y, Z) = P(X|Z)$

- Graphical models express sets of conditional independence assumptions via graph structure
- The graph structure plus associated parameters define joint probability distribution over the set of variables

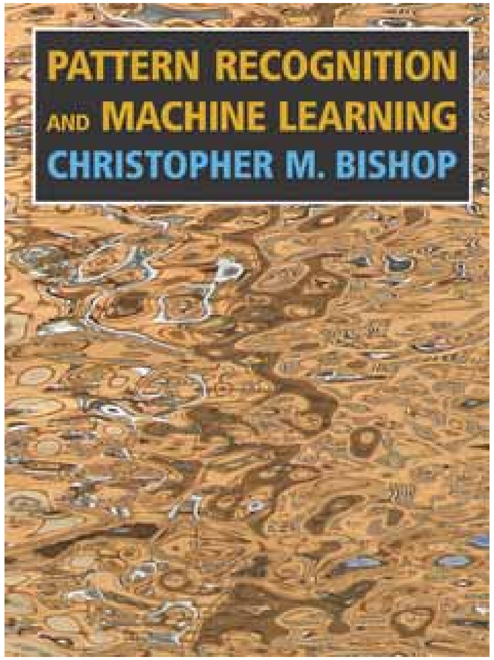
Where do the data come from?

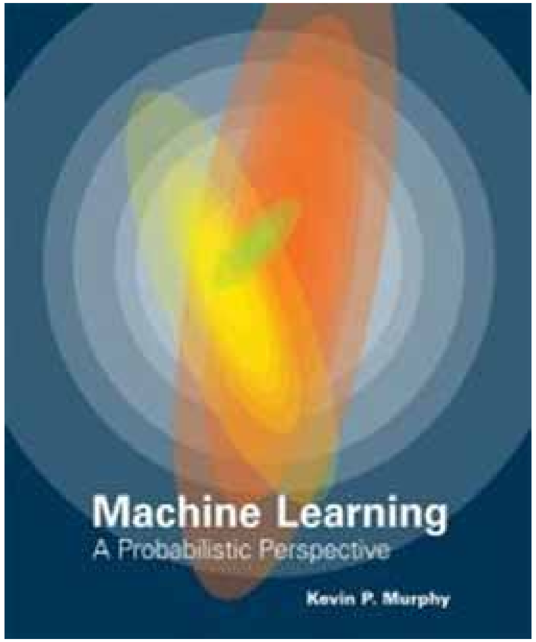


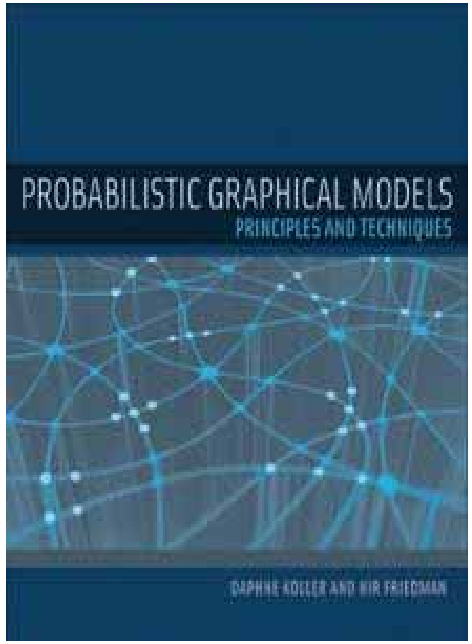
http://www.ebi.ac.uk/intact/

- Medicine is an extremely complex application domain – dealing most of the time with uncertainties -> **probable information!**
- When we have big data but little knowledge automatic ML can help to gain insight:
- **Structure learning and prediction in large-scale biomedical networks with probabilistic graphical models**
- If we have little data and deal with NP-hard problems we still need the human-in-the-loop

Basics and Background reading





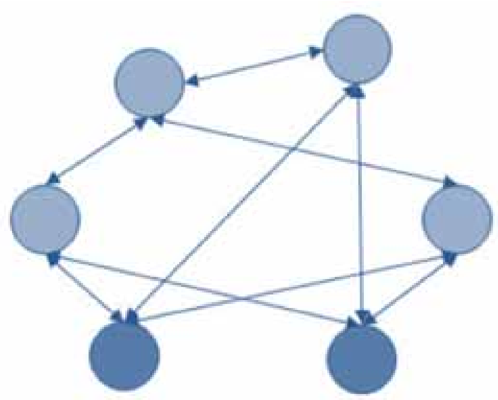


Bishop, C. M. 2007. Pattern Recognition and Machine Learning, Heidelberg, Springer. Chapter 8 on graphical models openly available: http://research.microsoft.com/en-us/um/people/cmbishop/prml/

Murphy, K. P. 2012. Machine learning: a probabilistic perspective, MIT press. Chapter 26 (pp. 907) – Graphical model structure learning

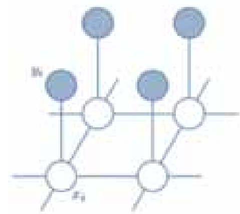
Koller, D. & Friedman, N. 2009. Probabilistic graphical models: principles and techniques, MIT press.

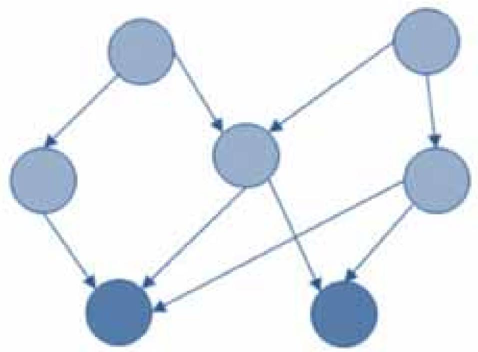
Three types of Probabilistic Graphical Models



Undirected: Markov random fields, useful e.g. for computer vision (Details: Murphy 19)

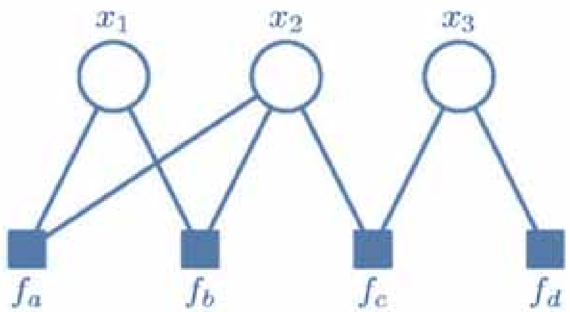
$$P(\boldsymbol{X}) = \frac{1}{Z} \exp\left(\sum_{ij} W_{ij}\, x_i x_j + \sum_i x_i b_i \right)$$





Directed: Bayes Nets, useful for designing models (Details: Murphy 10)

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k | \mathrm{pa}_k)$$



Factored: useful for inference/learning

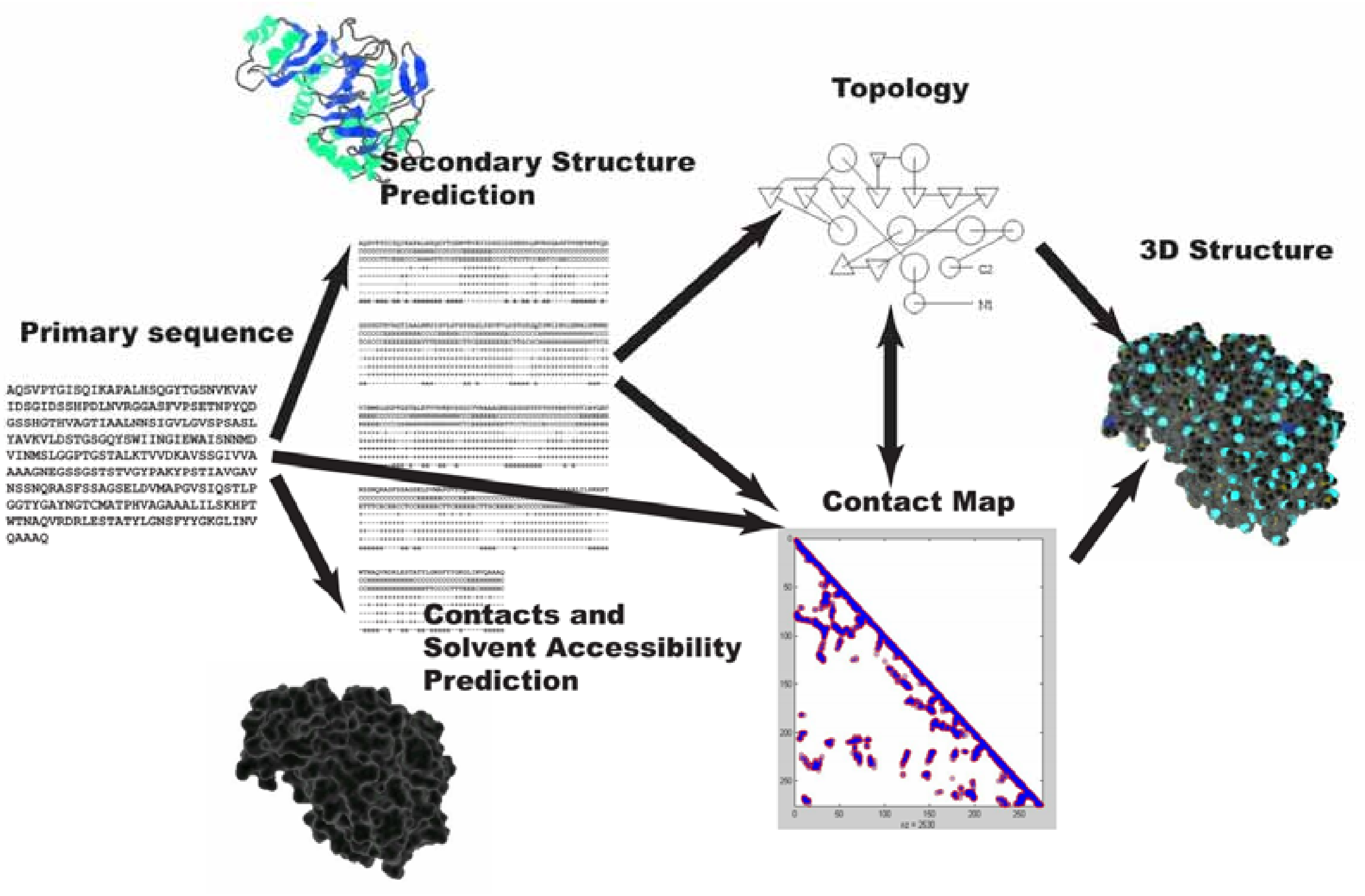
$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

- What is the advantage of factor graphs?

	Dependency	Efficient Inference	Usage
Bayesian Networks	Yes	Somewhat	Ancestral Generative Process
Markov Networks	Yes	No	Local Couplings and Potentials
Factor Graphs	No	Yes	Efficient, distributed inference

Table credit to Ralf Herbrich, Amazon

From structure to function prediction



Baldi, P. & Pollastri, G. 2003. The principled design of large-scale recursive neural network architectures--dag-rnns and the protein structure prediction problem. The Journal of Machine Learning Research, 4, 575-602.

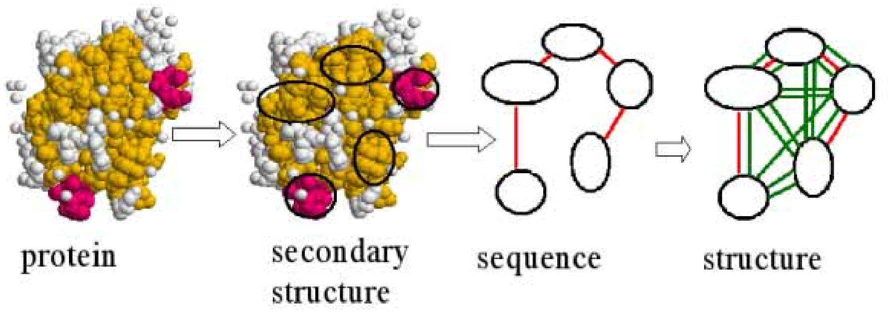
Protein Network Inference

- Hypothesis: most biological functions involve the interactions between many proteins, and the complexity of living systems arises as a result of such interactions.
- In this context, the problem of inferring a global protein network for a given organism,
- - using all (genomic) data of the organism,
- is one of the main challenges in computational biology

Yamanishi, Y., Vert, J.-P. & Kanehisa, M. 2004. Protein network inference from multiple genomic data: a supervised approach. Bioinformatics, 20, (suppl 1), i363-i370.

Problem: Is Graph Isomorphism NP-complete ?

Borgwardt, K. M., Ong, C. S., Schönauer, S., Vishwanathan, S., Smola, A. J. & Kriegel, H.-P. 2005. Protein function prediction via graph kernels. Bioinformatics, 21, (suppl 1), i47-i56.



- Important for health informatics: Discovering relationships between biological components
- Unsolved problem in computer science:
- Can the graph isomorphism problem be solved in polynomial time?
  - So far, no polynomial time algorithm is known.
  - It is also not known if it is NP-complete
  - We know that subgraph-isomorphism is NP-complete

Example: Protein Network Inference

BIOINFORMATICS

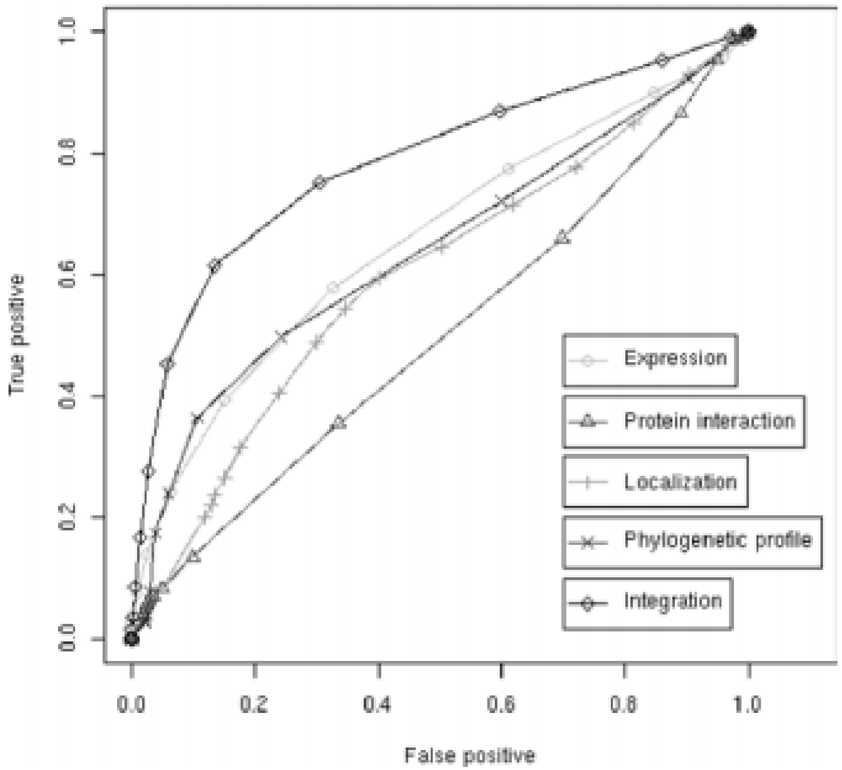
Vol. 20 Suppl. 1 2004, pages i363–i370
DOI: 10.1093/bioinformatics/bth910

Protein network inference from multiple genomic data: a supervised approach

Y. Yamanishi[1,*], J.-P. Vert[2] and M. Kanehisa[1]

[1]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan and [2]Computational Biology group, Ecole des Mines de Paris, 35 rue Saint-Honoré, 77305 Fontainebleau cedex, France

K_{exp} (Expression)
K_{ppi} (Protein interaction)
K_{loc} (Localization)
K_{phy} (Phylogenetic profile)
$K_{\text{exp}} + K_{\text{ppi}} + K_{\text{loc}} + K_{\text{phy}}$ (Integration)



Example: Data fusion and Protein Annotation

BIOINFORMATICS

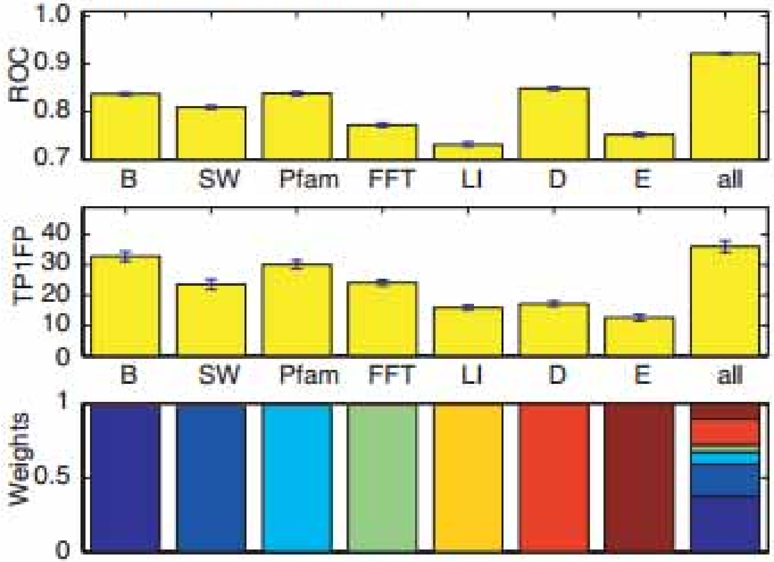
Vol. 20 no. 16 2004, pages 2626–2635
doi:10.1093/bioinformatics/bth294

A statistical framework for genomic data fusion

Gert R. G. Lanckriet[1], Tijl De Bie[3], Nello Cristianini[4], Michael I. Jordan[2] and William Stafford Noble[5,*]

[1]Department of Electrical Engineering and Computer Science, [2]Division of Computer Science, Department of Statistics, University of California, Berkeley 94720, USA, [3]Department of Electrical Engineering, ESAT-SCD, Katholieke Universiteit Leuven 3001, Belgium, [4]Department of Statistics, University of California, Davis 95618, USA and [5]Department of Genome Sciences, University of Washington, Seattle 98195, USA

Kernel	Data	Similarity measure
K_{SW}	protein sequences	Smith-Waterman
K_{B}	protein sequences	BLAST
K_{Pfam}	protein sequences	Pfam HMM
K_{FFT}	hydropathy profile	FFT
K_{LI}	protein interactions	linear kernel
K_{D}	protein interactions	diffusion kernel
K_{E}	gene expression	radial basis kernel
K_{RND}	random numbers	linear kernel



(B) Membrane proteins

Lanckriet, G. R., De Bie, T., Cristianini, N., Jordan, M. I. & Noble, W. S. 2004. A statistical framework for genomic data fusion. Bioinformatics, 20, (16), 2626-2635.

Bayesian Network (BN) - Definition

- is a **probabilistic model**, consisting of two parts:
- 1) a dependency structure and
- 2) local probability models.

$$p(x_1, \dots, x_n) = \prod_{i=1}^{n} p(x_i \mid Pa(x_i))$$

Where $Pa(x_i)$ are the parents of x_i

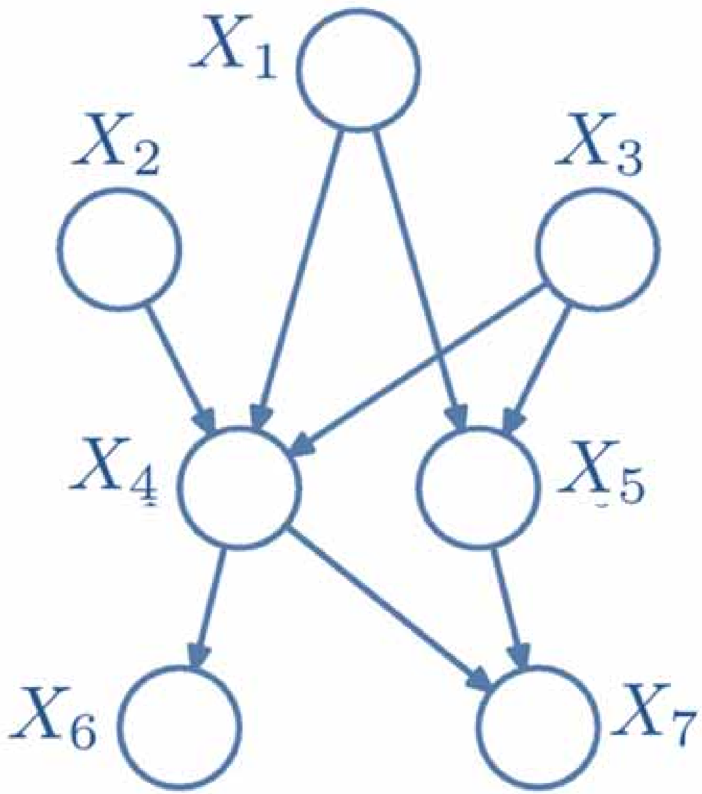
BN inherently model the **uncertainty in the data.** They are a successful marriage between probability theory and graph theory; allow to model a multidimensional probability distribution in a sparse way by searching independency relations in the data. Furthermore this model allows different strategies to integrate two data sources.

Pearl, J. (1988) *Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco, Morgan Kaufmann.*

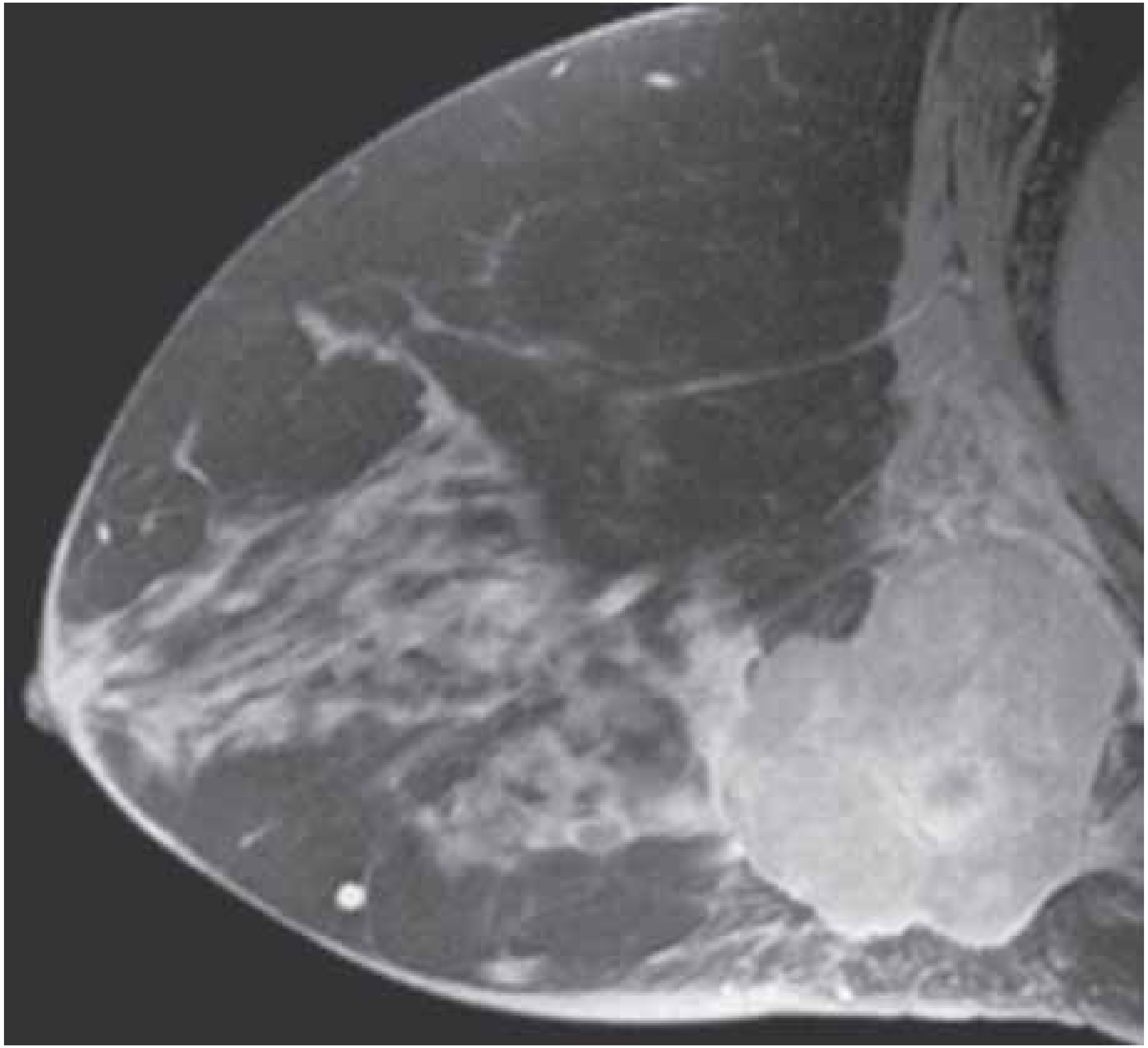
Example: Directed Bayesian Network with 7 nodes

$$p(X_1, \ldots, X_7) =$$

$$p(X_1)p(X_2)p(X_3)p(X_4|X_1, X_2, X_3) \cdot$$
$$p(X_5|X_1, X_3)p(X_6|X_4)p(X_7|X_4, X_5)$$



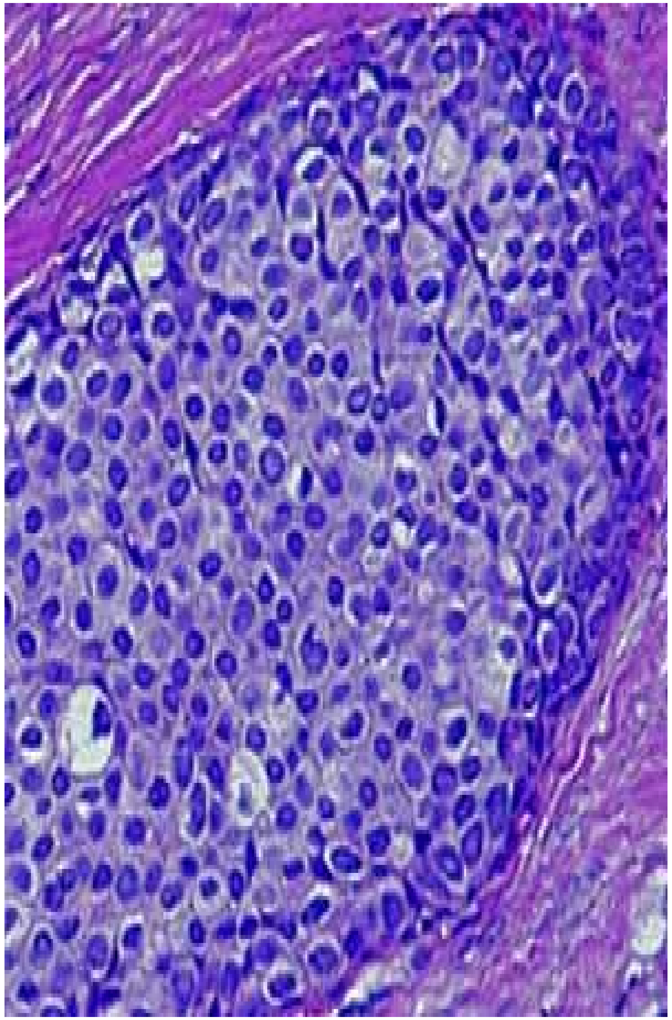
Clinical Case Example



Overmoyer, B. A., Lee, J. M. & Lerwill, M. F. (2011) Case 17-2011 A 49-Year-Old Woman with a Mass in the Breast and Overlying Skin Changes. *New England Journal of Medicine, 364, 23, 2246-2254.*

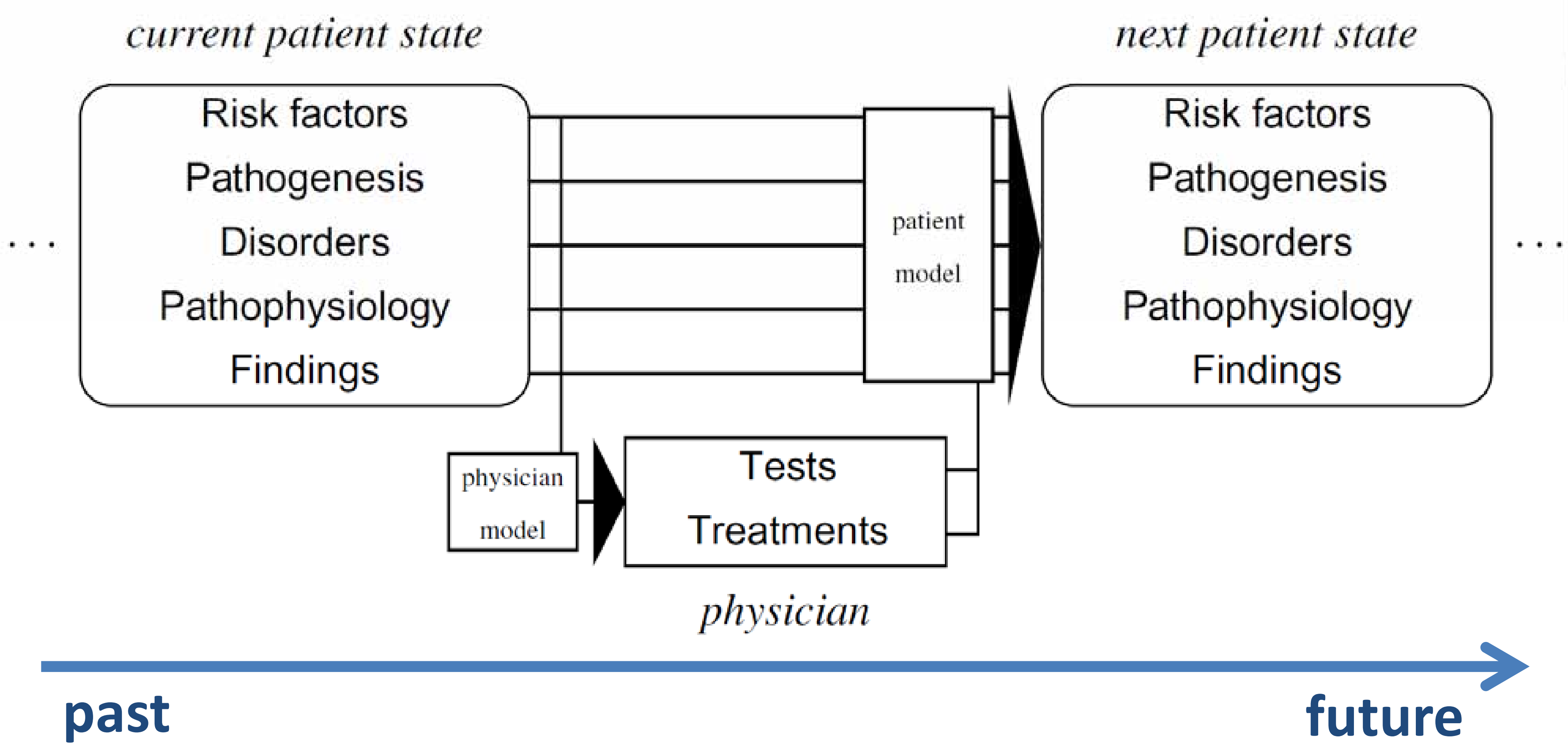
Important in Clinical practice -> prognosis !

- = the prediction of the future course of a disease conditional on the patient's history and a projected treatment strategy
- Danger: probable Information !
- Therefore valid prognostic models can be of great benefit for clinical decision making and of great value to the patient, e.g., for notification and quality of-life decisions



Knaus, W. A., Wagner, D. P. & Lynn, J. (1991) Short-term mortality predictions for critically ill hospitalized adults: science and ethics. *Science, 254, 5030, 389.*

Predicting the future on past data and present status



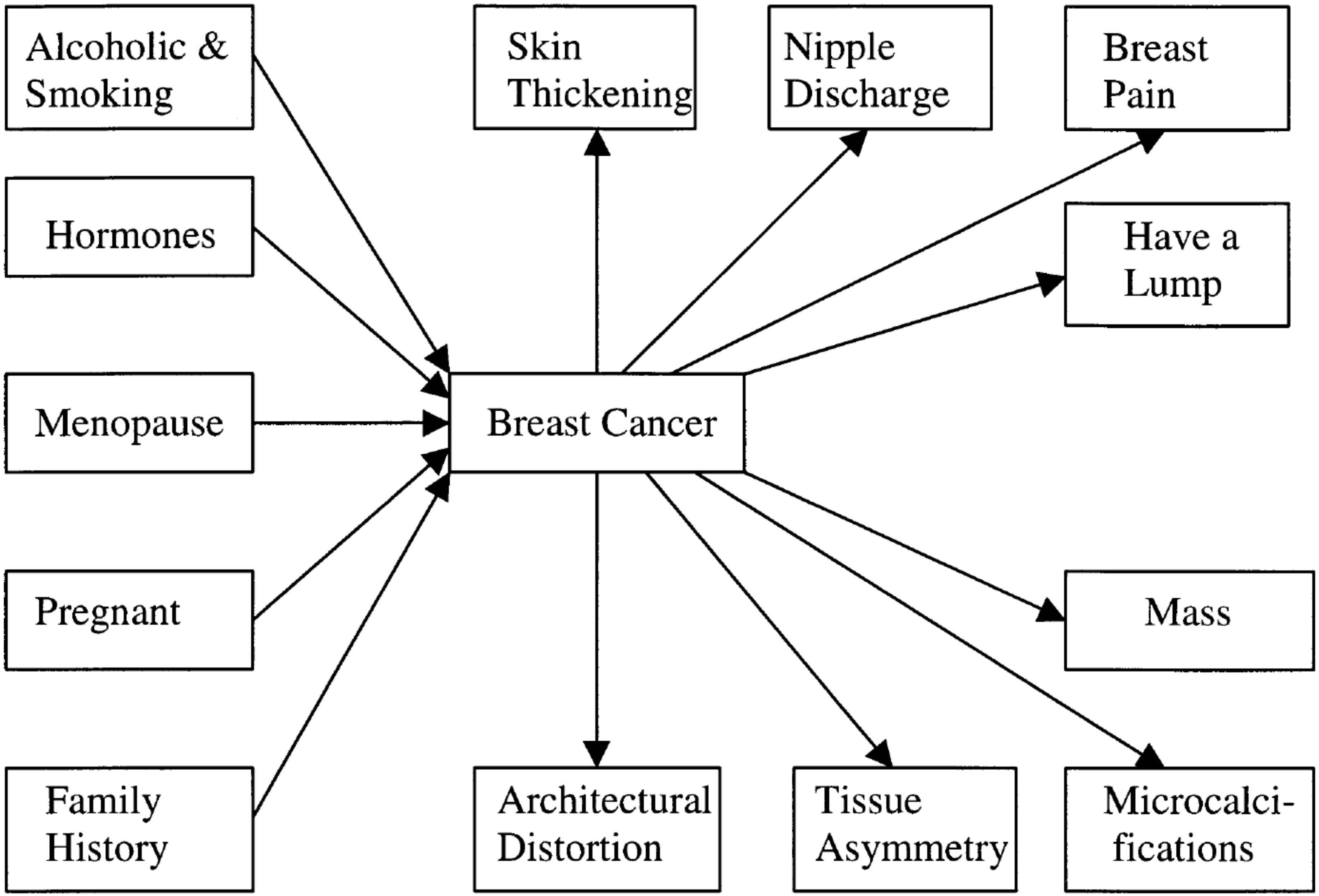
van Gerven, M. A. J., Taal, B. G. & Lucas, P. J. F. (2008) Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics, 41, 4, 515-529.*

Example: Breast cancer - Probability Table

Category	Node description	State description
Diagnosis	Breast cancer	Present, absent.
Clinical history	Habit of drinking alcoholic beverages and smoking	Yes, no.
	Taking female hormones	Yes, no.
	Have gone through menopause	Yes, no.
	Have ever been pregnant	Yes, no.
	Family member has breast cancer	Yes, no.
Physical findings	Nipple discharge	Yes, no.
	Skin thickening	Yes, no.
	Breast pain	Yes, no.
	Have a lump(s)	Yes, no.
Mammographic findings	Architectural distortion	Present, absent.
	Mass	Score from one to three, score from four to five, absent
	Microcalcification cluster	Score from one to three, score from four to five, absent
	Asymmetry	Present, absent.

Wang, X. H., et al. (1999) Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. *International Journal of Medical Informatics, 54, 2, 115-126.*

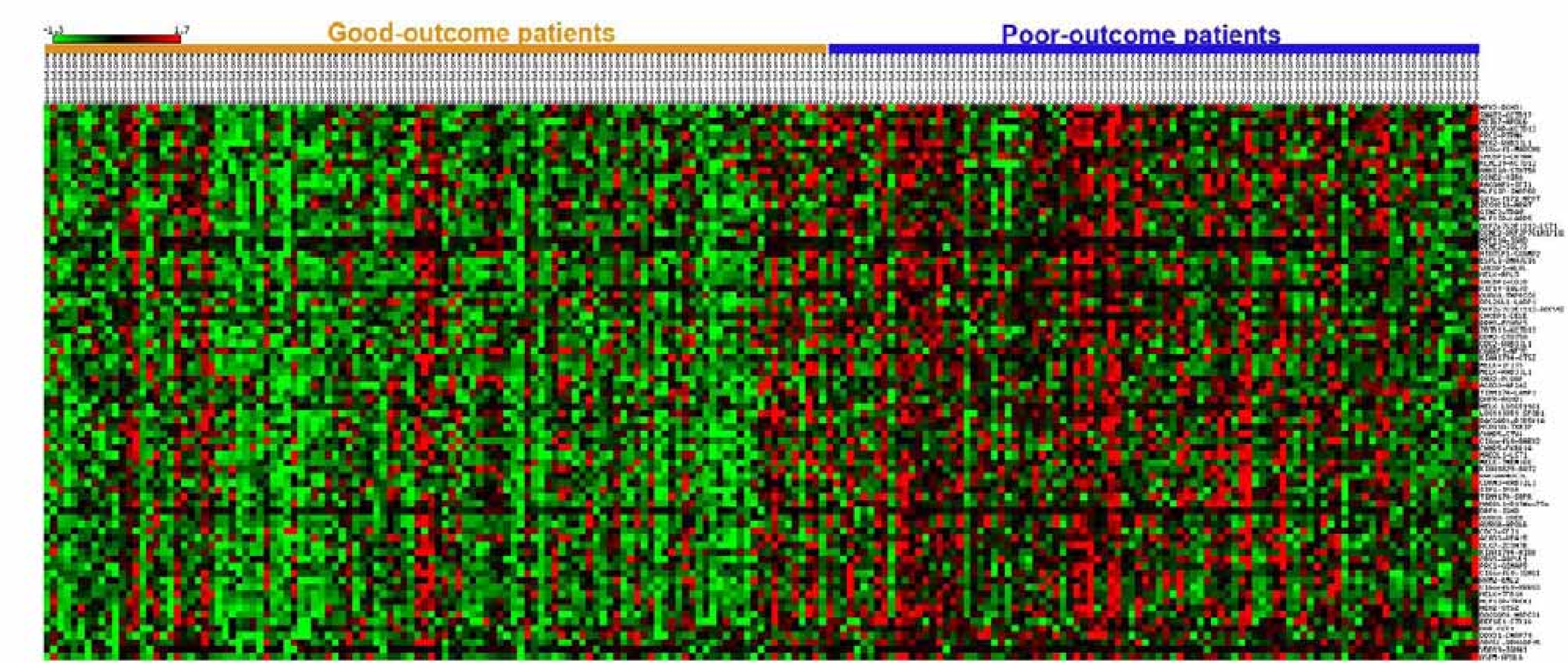
Breast cancer – big picture – state of 1999



Wang, X. H., et al. (1999) Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. *International Journal of Medical Informatics, 54, 2, 115-126.*

10 years later: Integration of microarray data

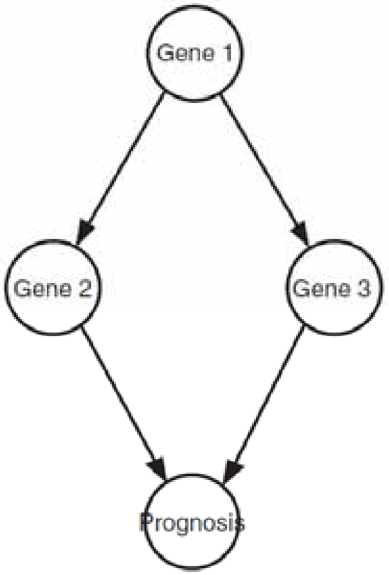
- Integrating microarray data from multiple studies to increase sample size;
- = approach to the development of more robust prognostic tests



Xu, L., Tan, A., Winslow, R. & Geman, D. (2008) Merging microarray data from separate breast cancer studies provides a robust prognostic test. *BMC Bioinformatics, 9, 1, 125-139.*

Example: BN with four binary variables

Gene 1	
P(on)	0.8
P(off)	0.2



Gene 2	Gene 1 on	Gene 1 off
P(on)	0.3	0.6
P(off)	0.7	0.4

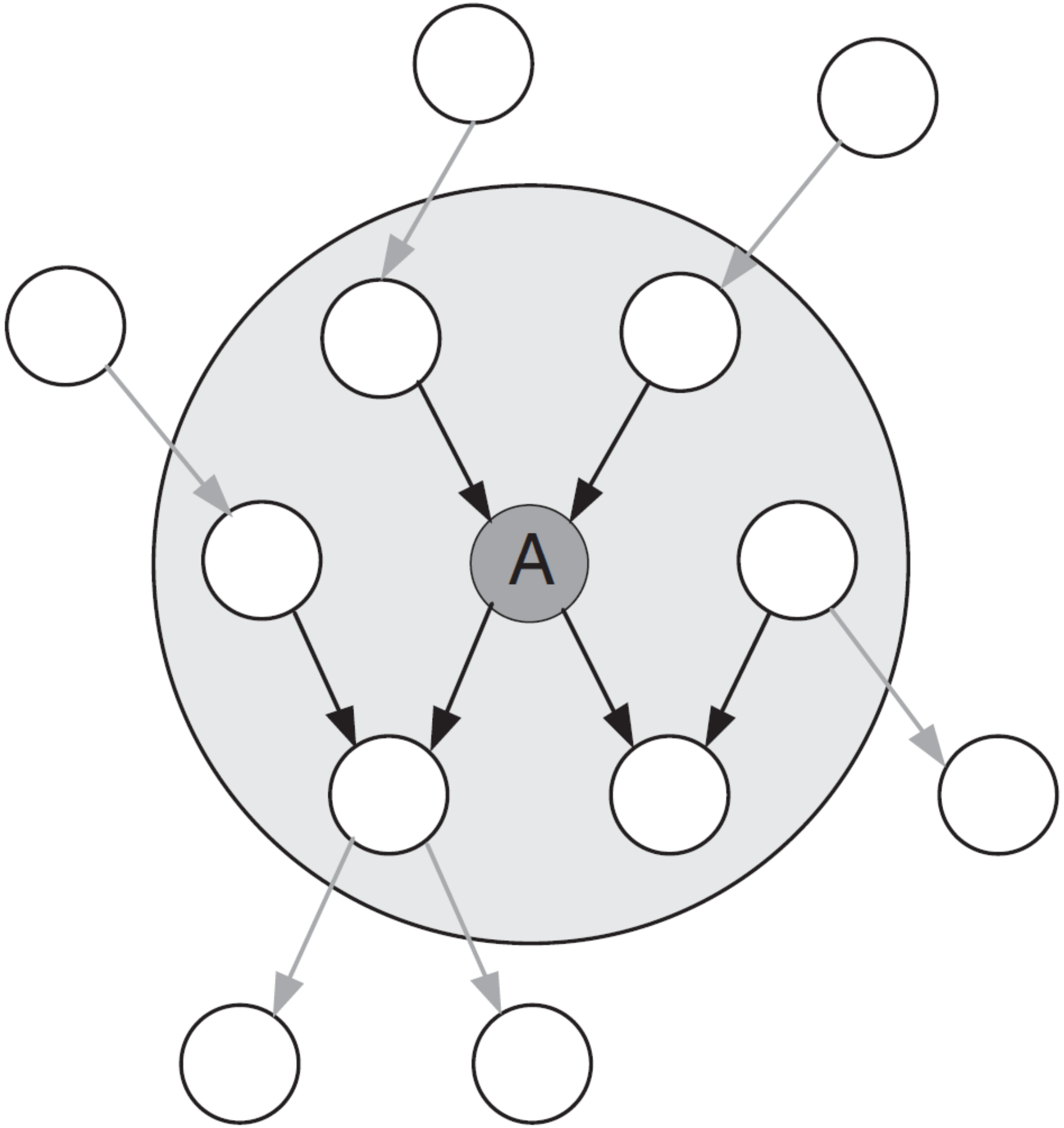
Gene 2	Gene 1 on	Gene 1 off
P(on)	0.3	0.6
P(off)	0.7	0.4

Prognosis	Gene 2 on Gene 3 on	Gene 2 on Gene 3 off	Gene 2 off Gene 2 on	Gene 2 off Gene 3 off
P(good)	0.6	0.1	0.9	0.5
P(poor)	0.4	0.9	0.1	0.5

Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y. & Moor, B. D. (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics, 22, 14, 184-190.*

Concept Markov-Blanket

Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y. & Moor, B. D. (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics, 22, 14, 184-190.*



- First the structure is learned using a search strategy.
- Since the number of possible structures increases super exponentially with the number of variables,
- the well-known greedy search algorithm K2 can be used in combination with the Bayesian Dirichlet (BD) scoring metric:

$$p(S|D) \propto p(S) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \left[\frac{\Gamma(N'_{ij})}{\Gamma(N'_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N'_{ijk} + N_{ijk})}{\Gamma(N'_{ijk})} \right]$$

N_{ijk} … number of cases in the data set D
having variable i in state k associated with the j-th instantiation of its parents in current structure S.
n is the total number of variables.

Dependency Structure – first step (2/2)

- Next, N_{ij} is calculated by summing over all states of a variable:
- $N_{ij} = \sum_{k=1}^{r_i} N_{ijk} \cdot N'_{ijk}$ and N'_{ij} have similar meanings but refer to prior knowledge for the parameters.
- When no knowledge is available they are estimated using $N_{ijk} = N/(r_i q_i)$
- with N the equivalent sample size,
- r_i the number of states of variable i and
- q_i the number of instantiations of the parents of variable i.
- $\Gamma(.)$ corresponds to the gamma distribution.
- Finally $p(S)$ is the prior probability of the structure.
- $p(S)$ is calculated by:
- $p(S) = \prod_{i=1}^{n} \prod_{l_i=1}^{p_i} p(l_i \longrightarrow x_i) \prod_{m_i=1}^{o_i} p(m_i x_i)$
- with p_i the number of parents of variable x_i and o_i all the variables that are not a parent of x_i.
- Next, $p(a \longrightarrow b)$ is the probability that there is an edge from a to b while $p(ab)$ is the inverse, i.e. the probability that there is no edge from a to b

Parameter learning -> second step

- Estimating the parameters of the local probability models corresponding with the dependency structure.
- CPTs are used to model these local probability models.
- For each variable and instantiation of its parents there exists a CPT that consists of a set of parameters.
- Each set of parameters was given a uniform Dirichlet prior:

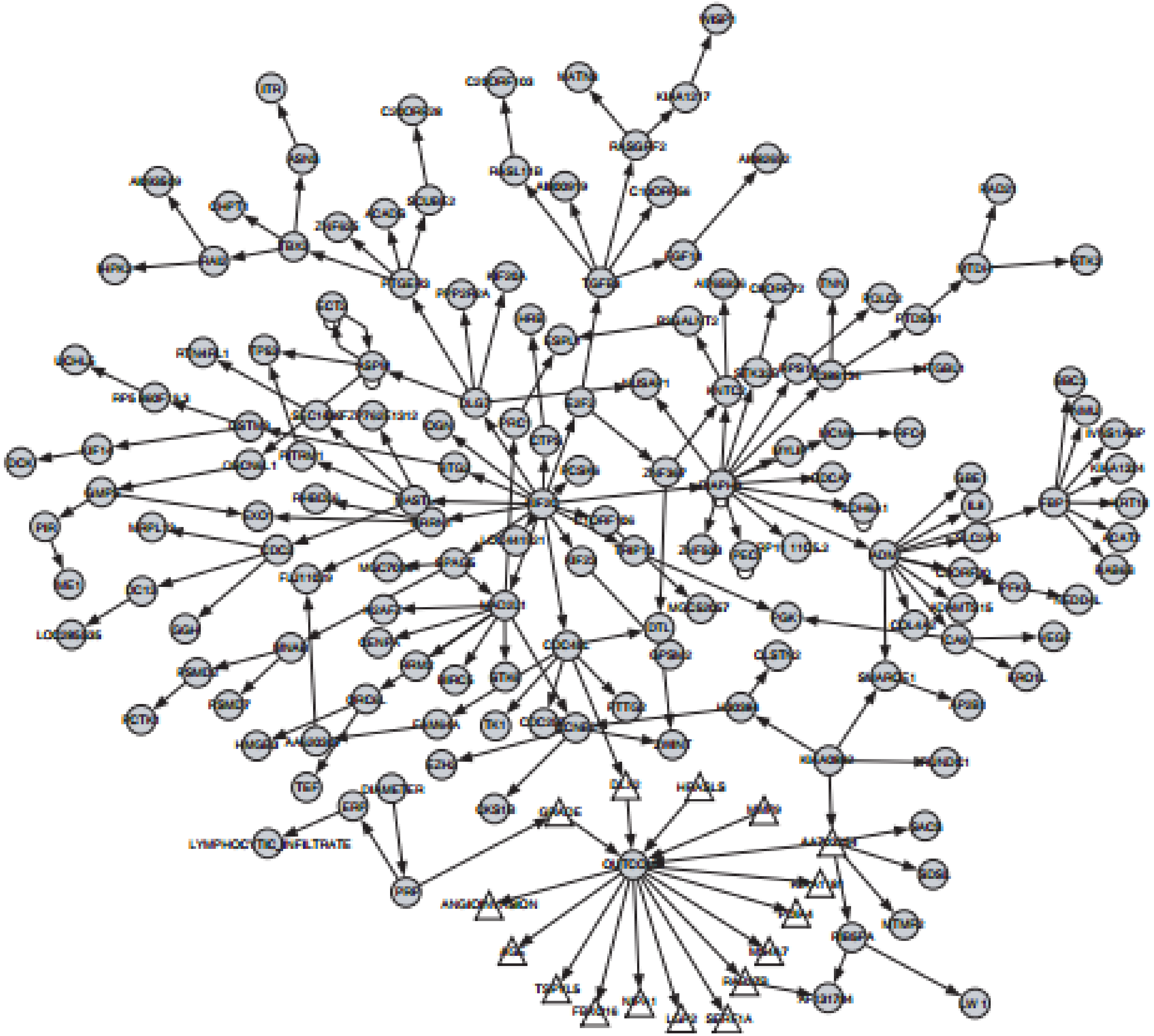
$$\mathrm{p}(\theta_{ij}|S) = Dir(\theta_{ij}|N'_{ij1}, \dots, N'_{ijk}, \dots, N'_{ijr_i})$$

Note: With θ_{ij} a parameter set where i refers to the variable and j to the j-th instantiation of the parents in the current structure. θ_{ij} contains a probability for every value of the variable x_i given the current instantiation of the parents. Dir corresponds to the Dirichlet distribution with $(N'_{ij1}, \dots, N'_{ijr_i})$ as parameters of this Dirichlet distribution. Parameter learning then consists of updating these Dirichlet priors with data. This is straightforward because the multinomial distribution that is used to model the data, and the Dirichlet distribution that models the prior, are conjugate distributions. This results in a Dirichlet posterior over the parameter set:

$$\mathrm{p}(\theta_{ij}|D, S) = Dir(\theta_{ij}|N'_{ij1} + N_{ij1}, \dots, N'_{ijk} + N_{ijk}, \dots, N'_{ijr_i} + N_{ijr_i})$$

with N_{ijk} defined as before.

Predicting the prognosis of breast cancer (integrated a.)



Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y. & Moor, B. D. (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics, 22, 14, 184-190.*

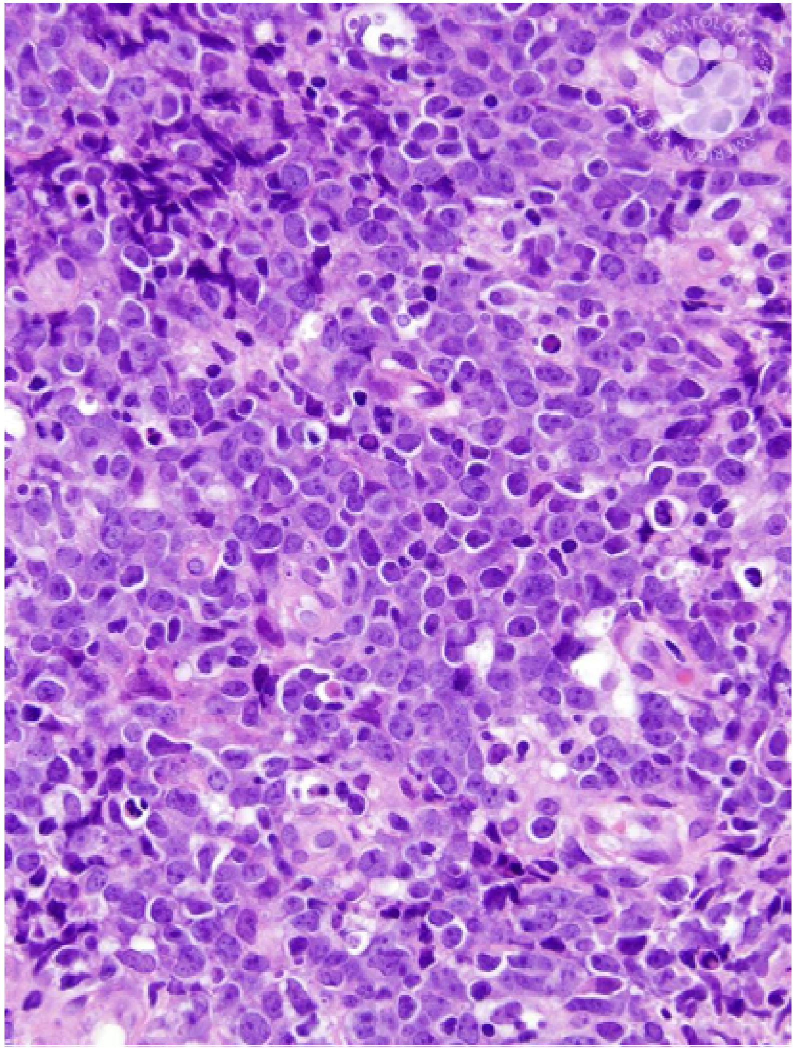
Inference in Bayes Nets is intractable (NP-complete!)

- For certain cases it is tractable if:
  - Just one variable is unobserved
  - We have singly connected graphs (no undirected loops -> belief propagation)
  - Assigning probability to fully observed set of variables
- Possibility: Monte Carlo Methods (generate many samples according to the Bayes Net distribution and then count the results)
- Otherwise: approximate solutions, NOTE:

Sometimes it is better to have an approximate solution to a complex problem – than a perfect solution to a simplified problem

Example: Lymphoma is the most common blood cancer

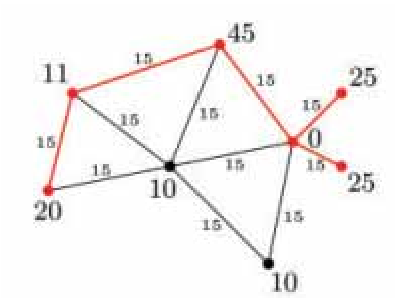
The two main forms of lymphoma are Hodgkin lymphoma and non-Hodgkin lymphoma (NHL). Lymphoma occurs when cells of the immune system called lymphocytes, a type of white blood cell, grow and multiply uncontrollably. Cancerous lymphocytes can travel to many parts of the body, including the lymph nodes, spleen, bone marrow, blood, or other organs, and form a mass called a tumor. The body has two main types of lymphocytes that can develop into lymphomas: B-lymphocytes (B-cells) and T-lymphocytes (T-cells).

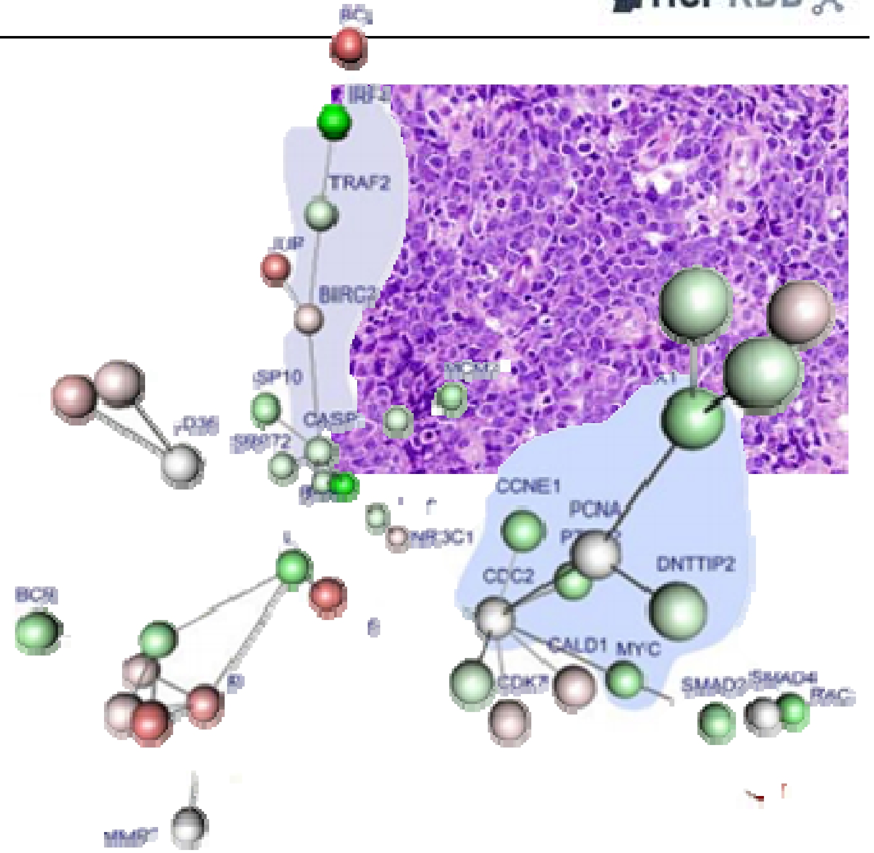


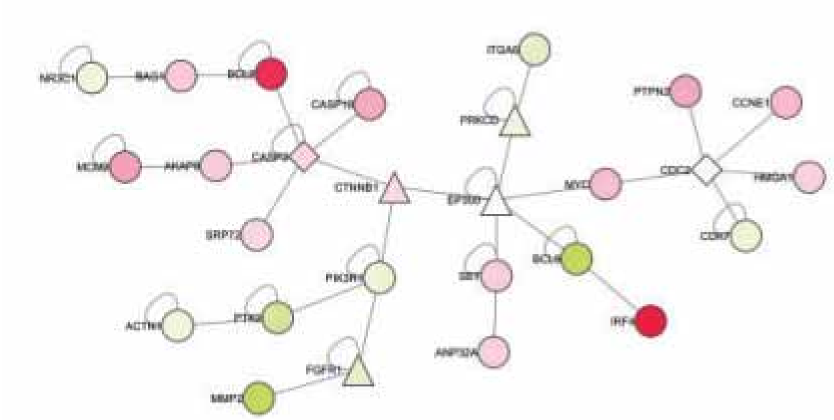
www.lymphoma.org

http://imagebank.hematology.org/

- Discover unexplored interactions in PPI-networks and gene regulatory networks
- Learn the structure
- Reconstruct the structure

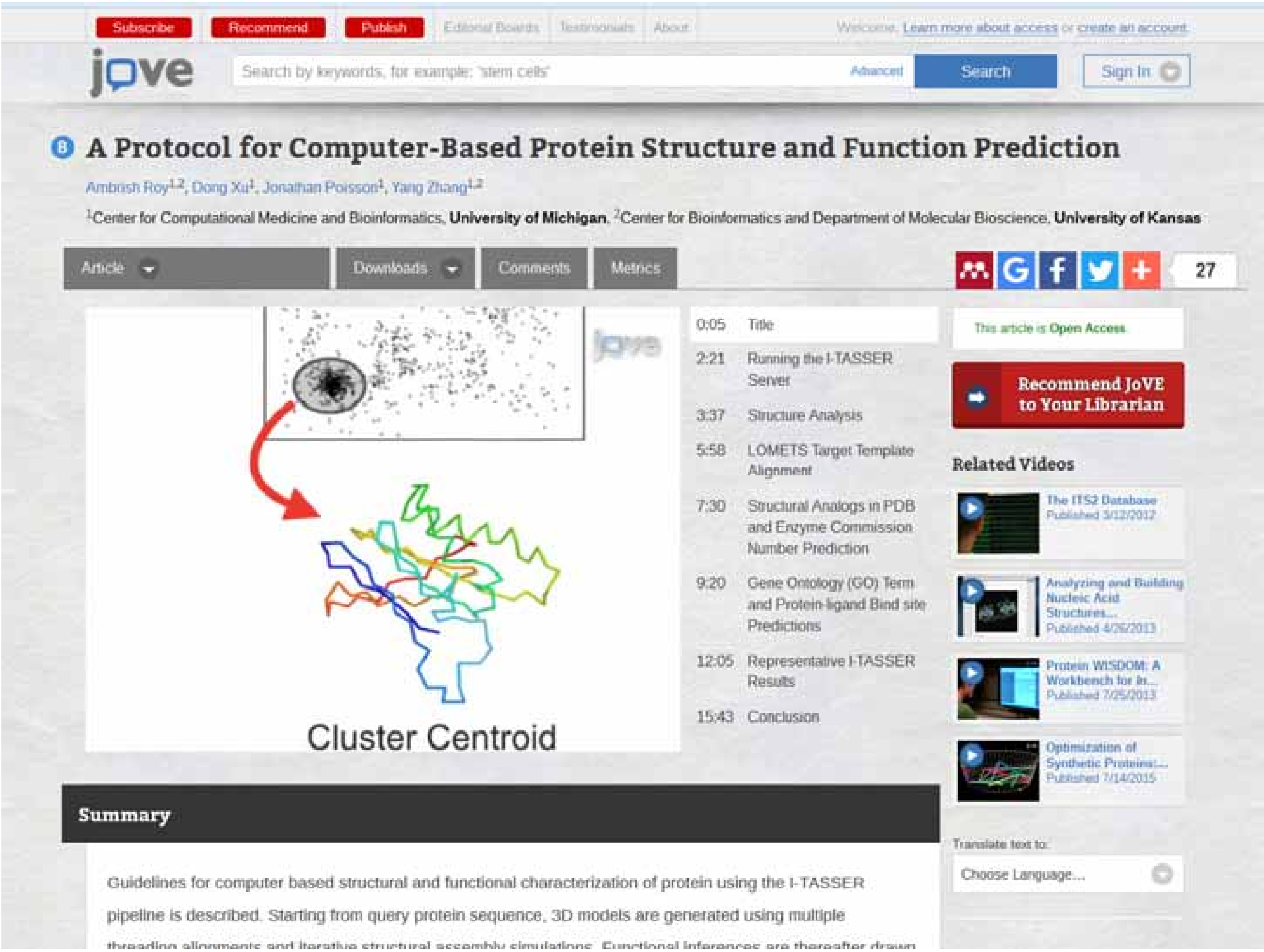






Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T. & Müller, T. 2008. Identifying functional modules in protein–protein interaction networks: an integrated exact approach. Bioinformatics, 24, (13), i223-i231.

From structure to function



http://www.jove.com/video/3259/a-protocol-for-computer-based-protein-structure-function

Interesting: Hubs tend to link to small degree nodes

Nodes: proteins

Links: physical interactions (binding)

Puzzling pattern:

Hubs tend to link to small degree nodes.

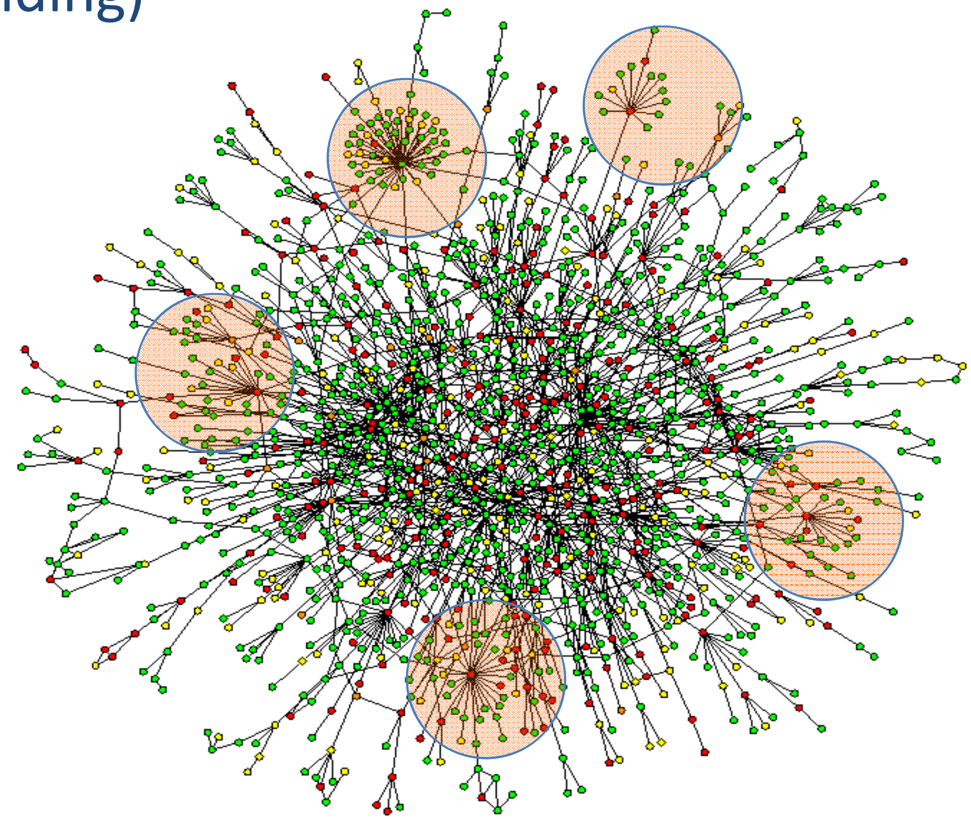
Why is this puzzling?

In a random network, the probability that a node with degree *k* links to a node with degree *k'* is:

$$p_{kk'} = \frac{kk'}{2L}$$

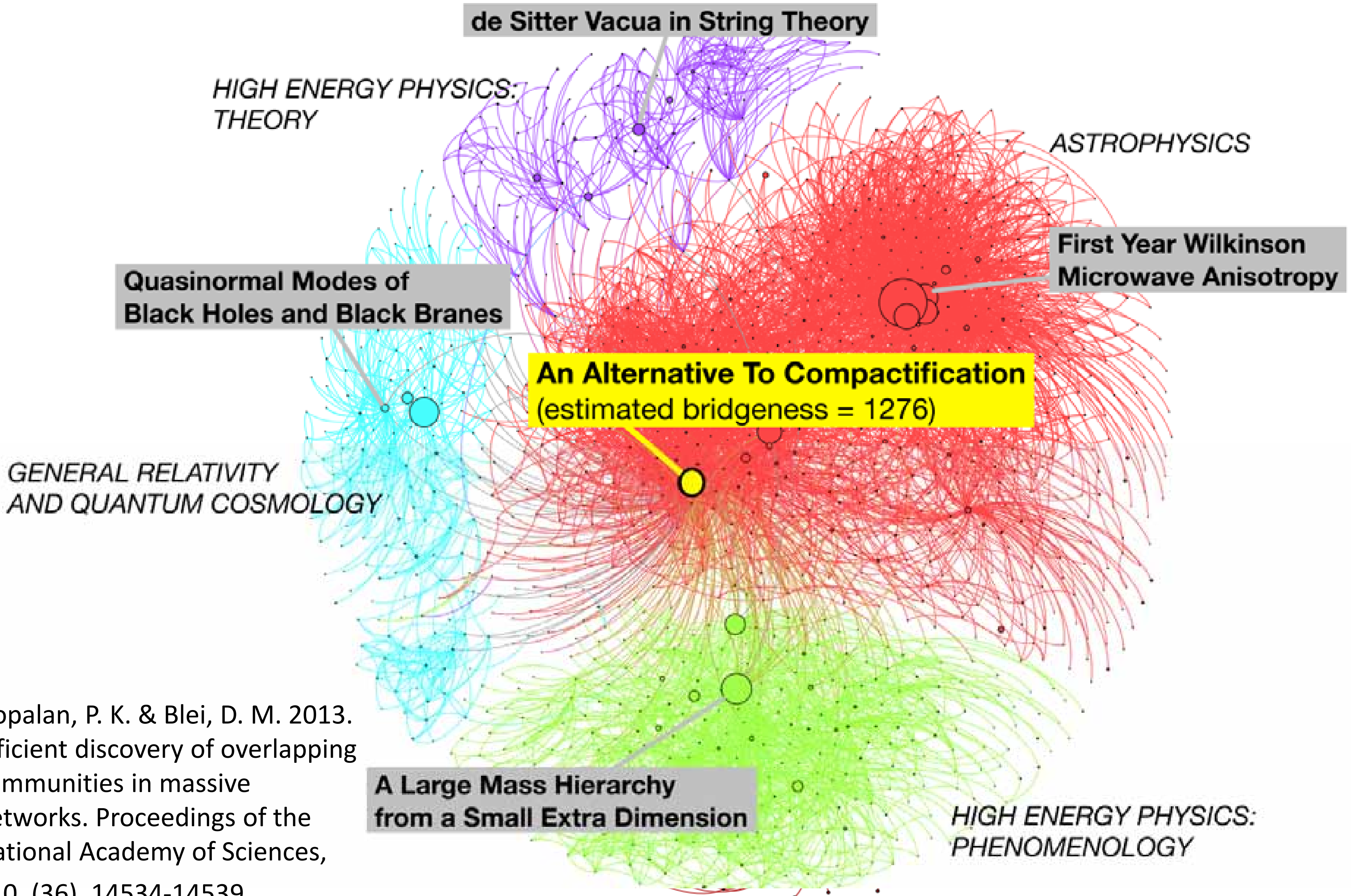
k≅50, k'=13, N=1,458, L=1746

$$p_{50,13} = 0.15 \qquad p_{2,1} = 0.0004$$



Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. 2001. Lethality and centrality in protein networks. Nature, 411, (6833), 41-42.

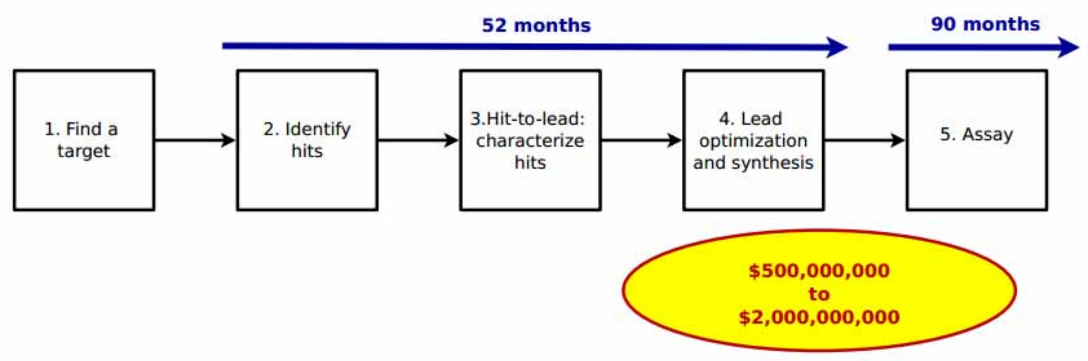
Example: Subgraph Discovery



Gopalan, P. K. & Blei, D. M. 2013. Efficient discovery of overlapping communities in massive networks. Proceedings of the National Academy of Sciences, 110, (36), 14534-14539.

Why do we want to apply ML to graphs

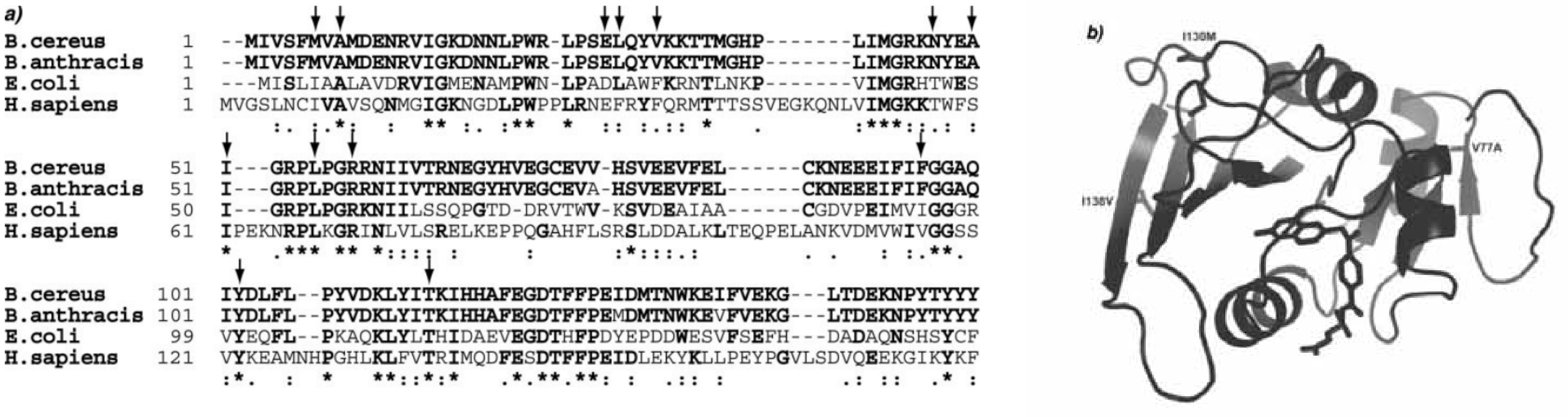
- A) Discovery of unexplored interactions
- B) Learning and Predicting the structure
- C) Reconstructing the structure
- Which joint probability distributions does a graphical model represent?
- How can we learn the parameters and structure of a graphical model?

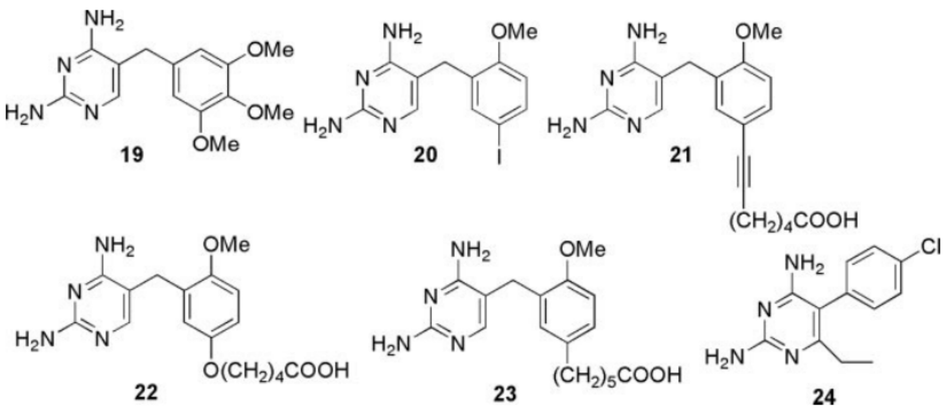


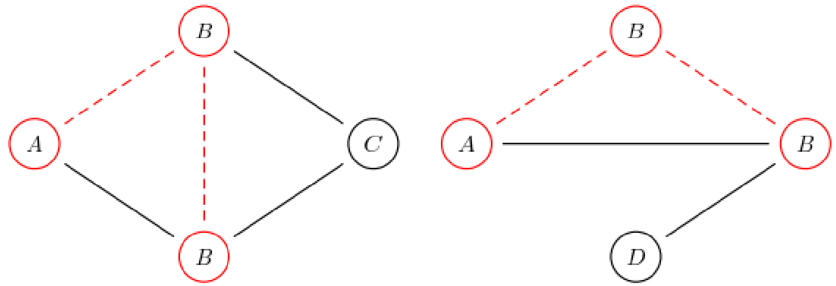
The chemical space

- 10^{60} possible small organic molecules
- 10^{22} stars in the observable universe

Example Question: Predicting Function from Structure







How similar are two graphs? How similar is their structure? How similar are their node and edge labels?

Joska, T. M. & Anderson, A. C. 2006. Structure-activity relationships of Bacillus cereus and Bacillus anthracis dihydrofolate reductase: toward the identification of new potent drug leads. Antimicrobial agents and chemotherapy, 50, 3435-3443.

- Remember: GM are a marriage between probability theory and graph theory and provide a tool for dealing with our two grand challenges in the biomedical domain:

Uncertainty and complexity

- The learning task is two-fold:
  1) Learning unknown probabilities
  2) Learning unknown structures

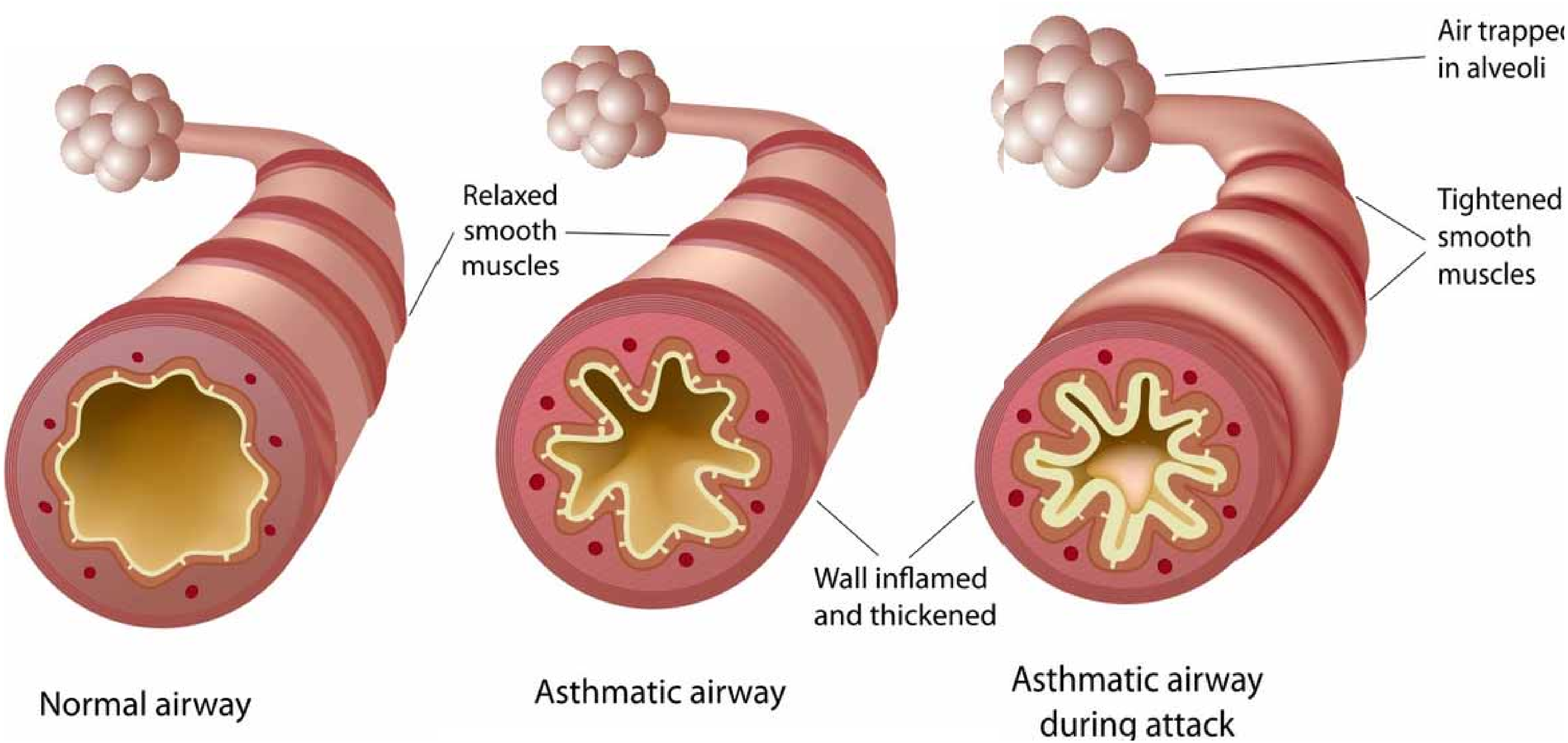
Jordan, M. I. 1998. Learning in graphical models, Springer

Learning the Structure of GM from data

1) Test if a distribution is decomposable with regard to a given graph.
   - This is the most direct approach. It is not bound to a graphical representation,
   - It can be carried out w.r.t. other representations of the set of subspaces to be used to compute the (candidate) decomposition of a given distribution.
2) Find a suitable graph by measuring the strength of dependences.
   - This is a heuristic, but often highly successful approach, which is based on the frequently valid assumption that in a conditional independence graph an attribute is more strongly dependent on adjacent attributes than on attributes that are not directly connected to them.
3) Find an independence map by conditional independence tests.
   - This approach exploits the theorems that connect conditional independence graphs and graphs that represent decompositions.
   - It has the advantage that a single conditional independence test, if it fails, can exclude several candidate graphs. Beware, because wrong test results can thus have severe consequences.

Borgelt, C., Steinbrecher, M. & Kruse, R. R. 2009. Graphical models: representations for learning, reasoning and data mining, John Wiley & Sons.

Who of you smokes?

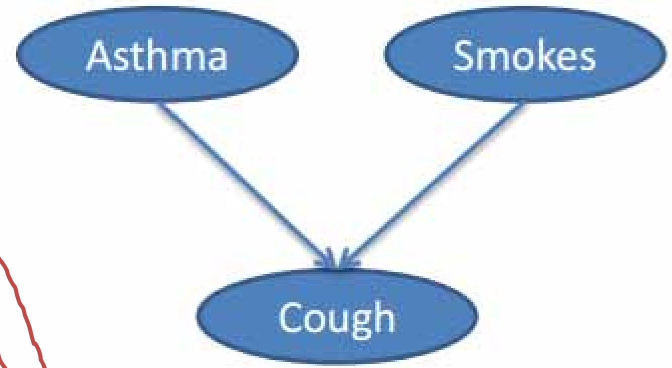


Beasley, R. 1998. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The Lancet, 351, (9111), 1225-1232, doi:http://dx.doi.org/10.1016/S0140-6736(97)07302-9.

Example for Graphical Model Learning

Bayesian Network

Patient	J46	Tussis	Smoker
Florian	1	1	0
Tamas	0	0	0
Matthias	1	0	0
Benjamin	0	1	1
Dimitrios	0	1	0
…			
…			

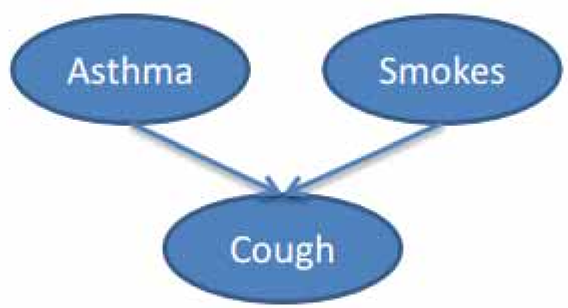


Florian	0	?	?

Florian	0	0.3	0.2

Rows are independent during learning and inference!

- Asthma can be hereditary
- Friends may have similar smoking habits
- Augmenting graphical model with relations between the entities – Markov Logic



2.1 Asthma ⇒ Cough

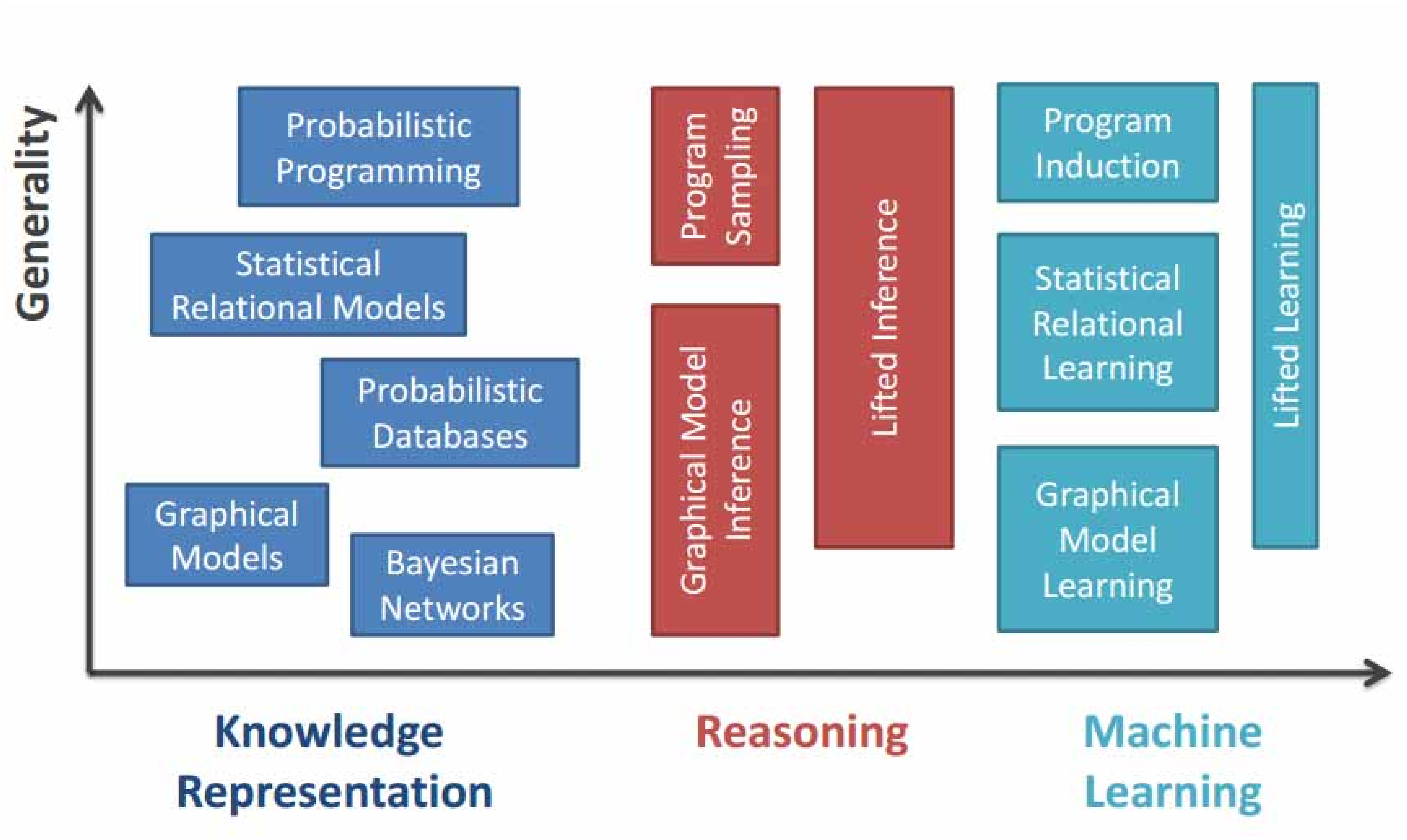
3.5 Smokes ⇒ Cough

2.1 Asthma(x) ⇒ Cough(x)

3.5 Smokes(x) ⇒ Cough(x)

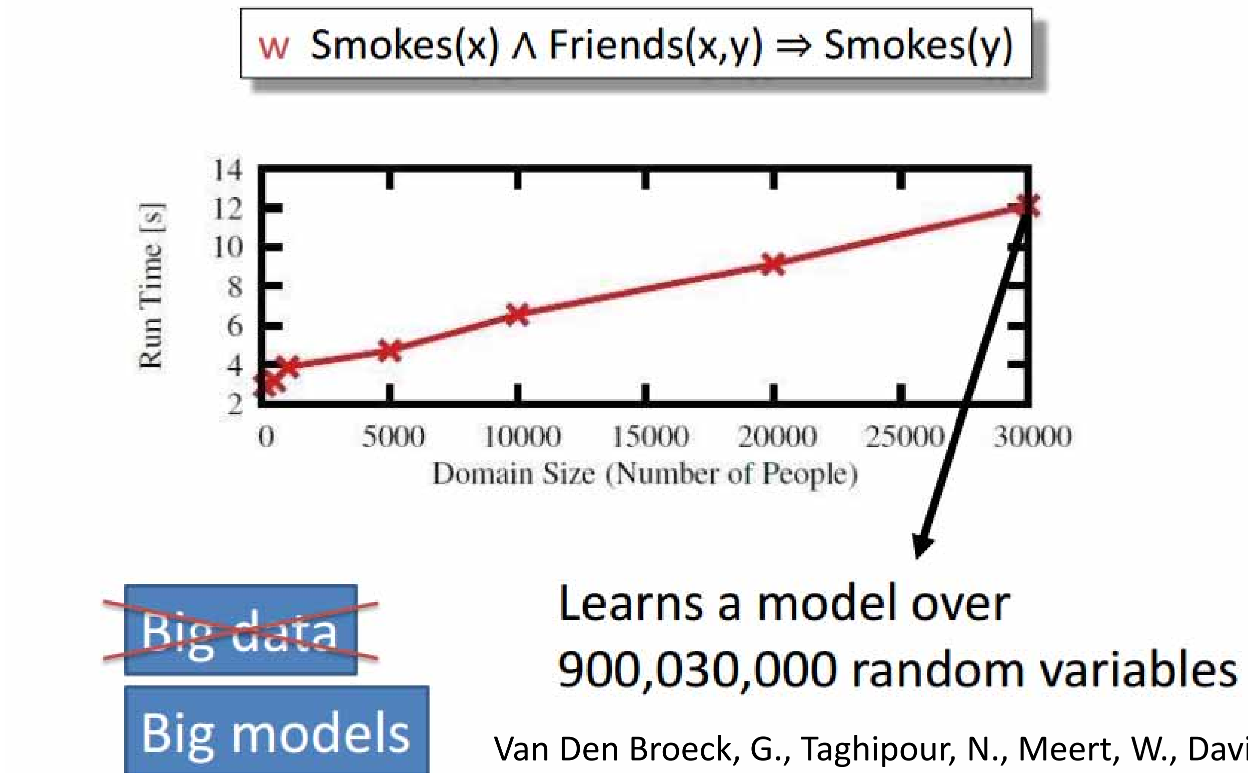
1.9 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

1.5 Asthma (x) ∧ Family(x,y) ⇒ Asthma (y)



Example for probabilistic rule learning, in which probabilistic rules are learned from probabilistic examples: The ProbFOIL+ Algorithm solves this problem by combining the principles of the rule learner FOIL with the probabilistic Prolog called ProbLog, see: De Raedt, L., Dries, A., Thon, I., Van Den Broeck, G. & Verbeke, M. 2015. Inducing probabilistic relational rules from probabilistic examples. International Joint Conference on Artificial Intelligence (IJCAI).

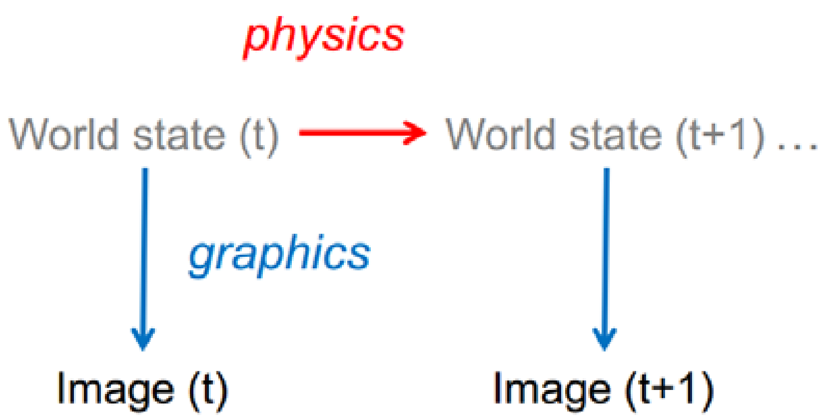
The future is in integrative ML, i.e. combining relational databases, ontologies and logic with probabilistic reasoning models and statistical learning – and algorithms that have good **scalability**

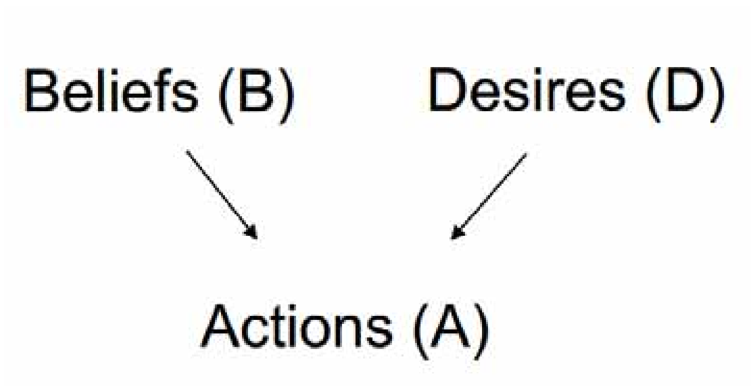


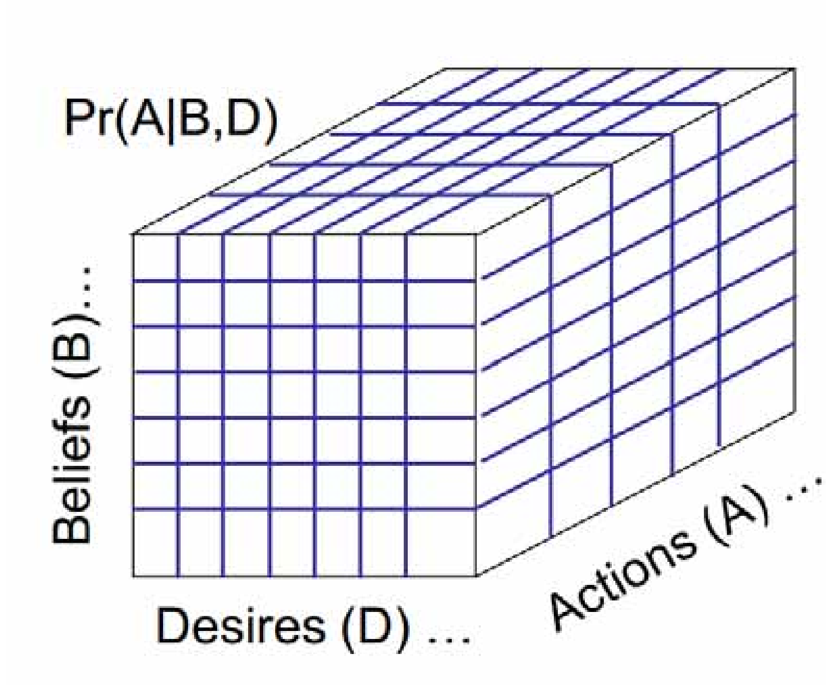
Van Den Broeck, G., Taghipour, N., Meert, W., Davis, J. & De Raedt, L. Lifted probabilistic inference by first-order knowledge compilation. Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three, 2011. AAAI Press, 2178-2185.

06 Probabilistic Programming

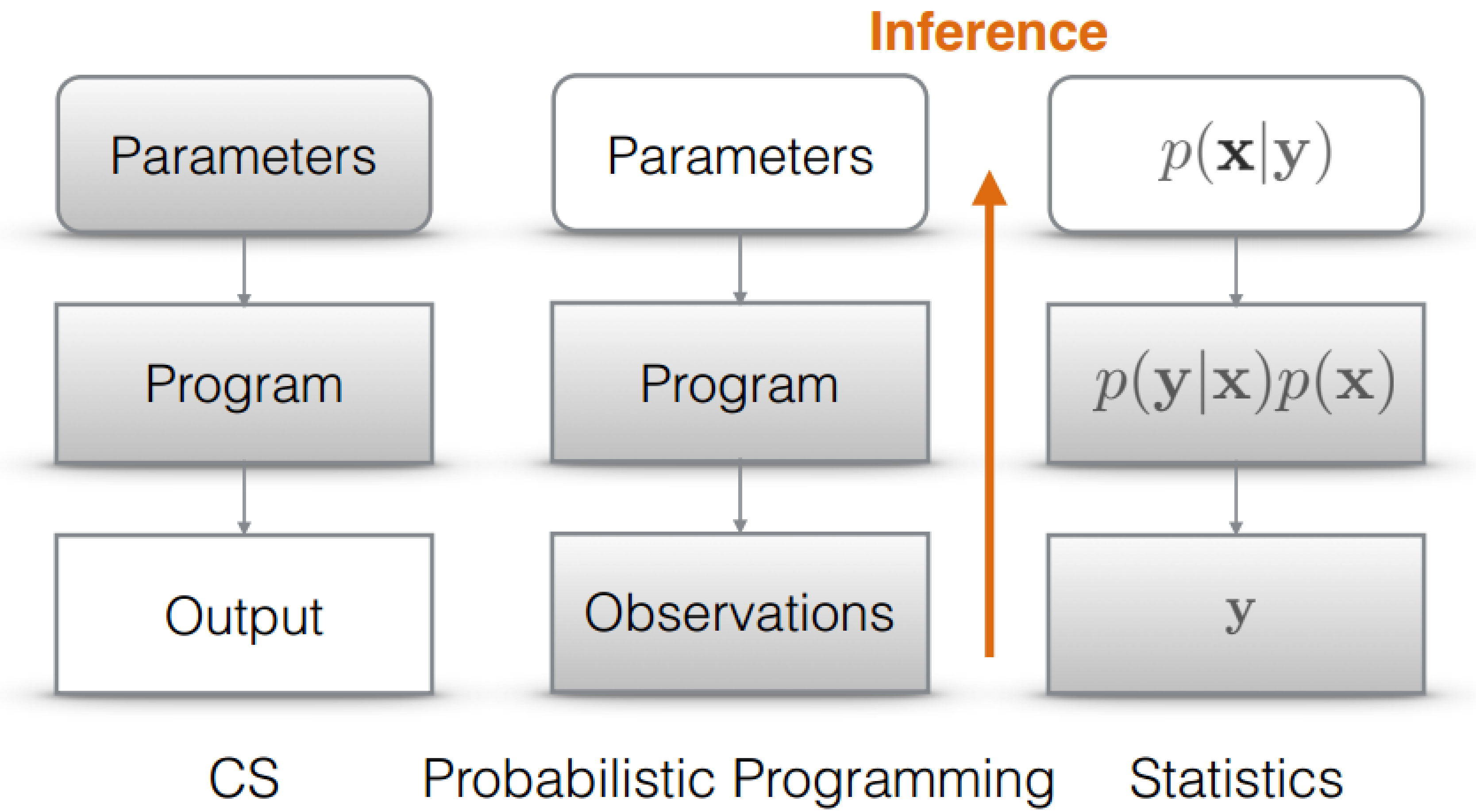
- Representatives for causal processes that are generative, relational, recursive, composable, and computationally universal







Intuition



Wood, F., Van De Meent, J.-W. & Mansinghka, V. A New Approach to Probabilistic Programming Inference. AISTATS, 2014. 1024-1032.

Defintions

- Probabilistic programs are usual functional or imperative programs with two added constructs:
  - (1) the ability to draw values at random from distributions, and
  - (2) the ability to condition values of variables in a program via observations.

  - Models from diverse application areas such as computer vision, coding theory, cryptographic protocols, biology and reliability analysis can be written as probabilistic programs. Probabilistic inference is the problem of computing an explicit representation of the probability distribution implicitly specified by a probabilistic program. Depending on the application, the desired output from inference may vary—we may want to estimate the expected value of some function f with respect to the distribution, or the mode of the distribution, or simply a set of samples drawn from the distribution.

Gordon, A. D., Henzinger, T. A., Nori, A. V. & Rajamani, S. K. Probabilistic programming. Proceedings of the on Future of Software Engineering, 2014. ACM, 167-181.

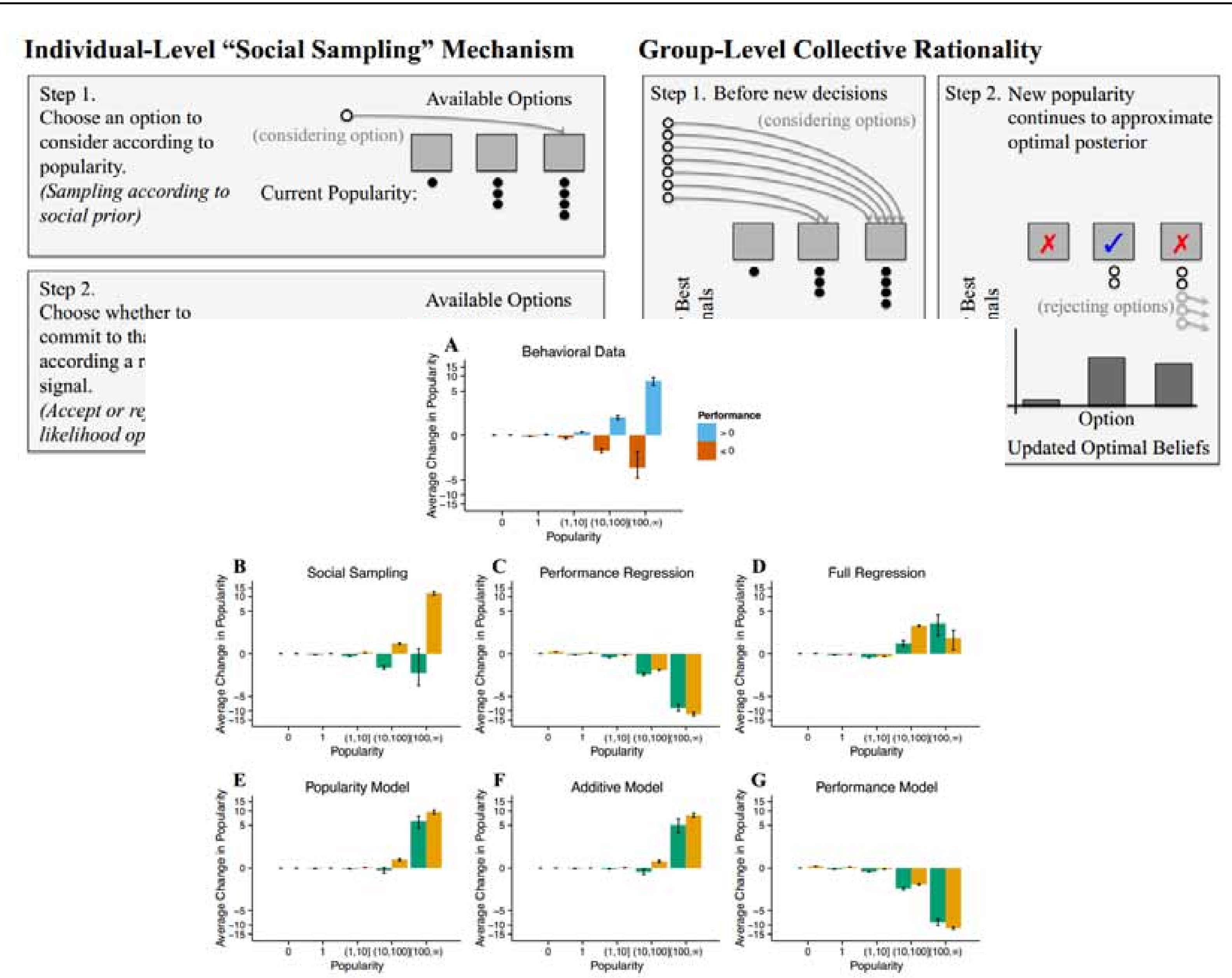
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

$\mathbf{x}$	$\mathbf{y}$
program source code	program output
scene description	image
policy and world	rewards
cognitive process	behavior
simulation	constraint

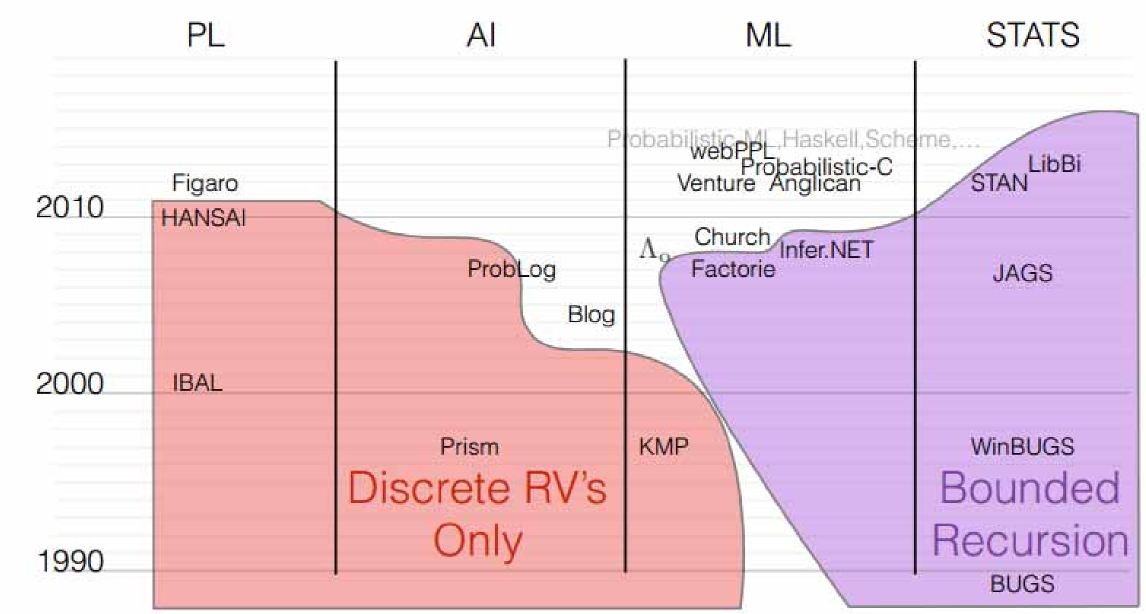
Tolpin, D., Van De Meent, J.-W. & Wood, F. Probabilistic programming in Anglican. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2015. Springer International Publishing, 308-311.

http://probcomp.csail.mit.edu/readings/ https://peerj.com/articles/cs-55/#p-5

Human collective Intelligence as distributed Bayesian I.



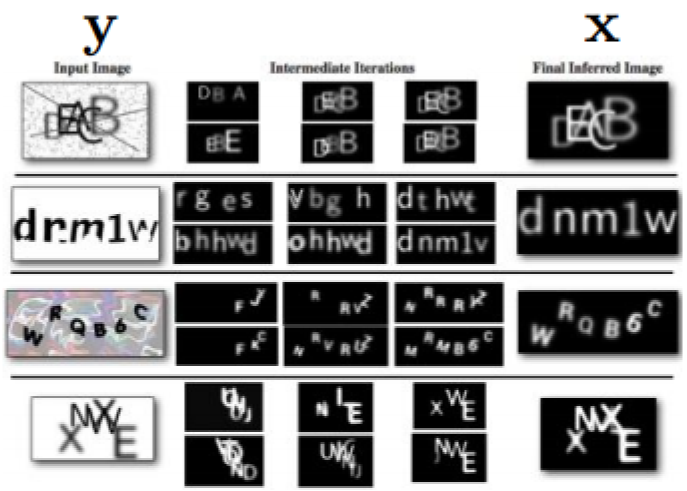
Historical Roots: Prolog



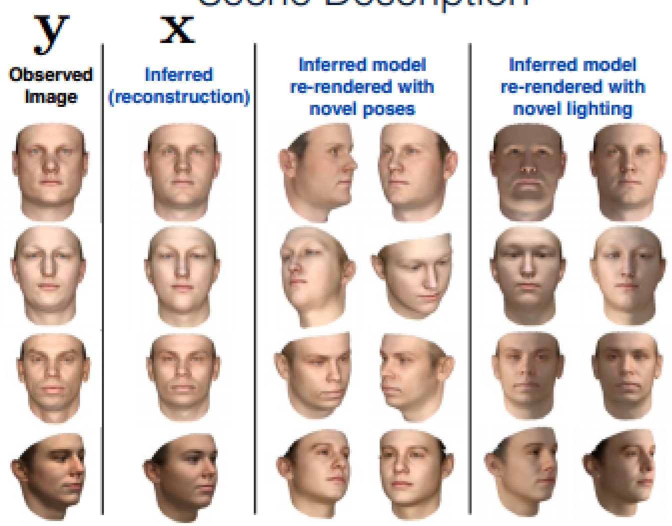
x	y
program source code	program output
scene description	image
policy and world	rewards
cognitive process	behavior
simulation	constraint

Applications: Perception and Inverse Graphics

Captcha Solving



Scene Description



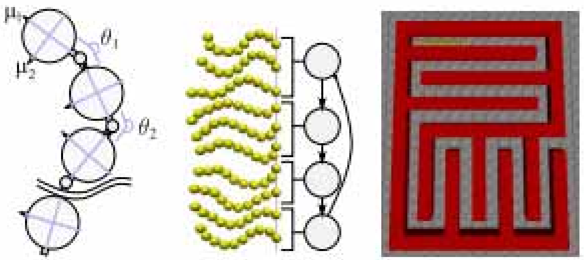
x	y
scene description	image

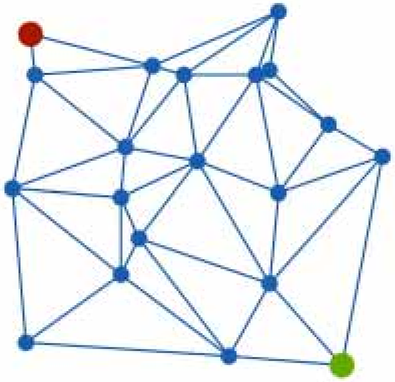
Mansinghka,, Kulkarni, Perov, and Tenenbaum. "Approximate Bayesian image interpretation using enerative probabilistic graphics programs." NIPS (2013).

Kulkarni, Kohli, Tenenbaum, Mansinghka "Picture: a probabilistic programming language for scene perception." CVPR (2015).

21

Application: Reinforcement Learning





x	**y**
policy and world	reward

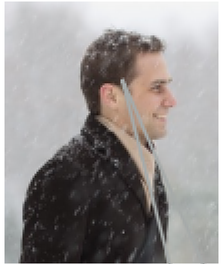
Wingate, Goodman, Roy, Kaelbling, and Tenenbaum. "Bayesian policy search with policy priors." (IJCAI), 2011.

van de Meent, Tolpin, Paige, and Wood. "Black-Box Policy Search with Probabilistic Programs." arXiv:1507.04635 (2015).

22

Reasoning about reasoning

Want to meet up but phones are dead…



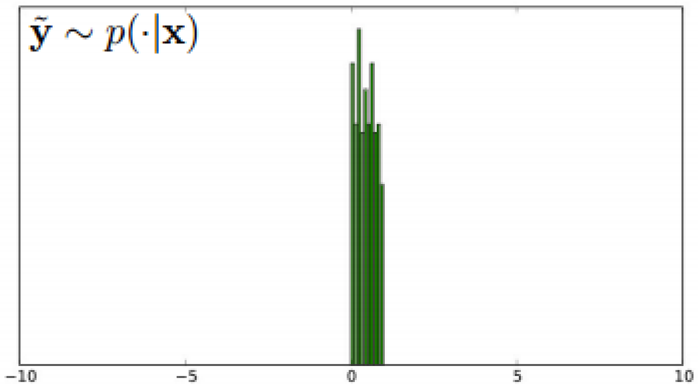
I prefer the pub.
Where will Noah go?
Simulate Noah:
Noah prefers pub
but will go wherever Andreas is
Simulate Noah simulating Andreas:
…
-> both go to pub

x	**y**
cognitive process	behavior

Stuhlmüller, and Goodman.
"Reasoning about reasoning by nested conditioning: Modeling theory of mind with probabilistic programs."
Cognitive Systems Research 28 (2014): 80-99.

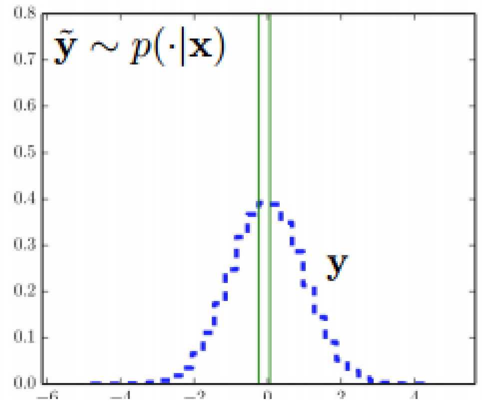
23

Program Induction




```
(lambda (stack-id) (safe-uc (* (if (< 0.0 (* (* (* -1.0 (begin (define
 G__1147 (safe-uc 1.0 1.0)) 0.0)) (* 0.0 (+ 0.0 (safe-uc (* (* (dec -2
.0) (safe-sqrt (begin (define G__1148 3.14159) (safe-log -1.0)))) 2.0)
 0.0)))) 1.0)) (+ (safe-div (begin (define G__1149 (* (+ 3.14159 -1.0)
 1.0)) 1.0) 0.0) (safe-log 1.0)) (safe-log -1.0)) (begin (define G__11
...
```

$\mathbf{x} \sim p(\mathbf{x})$



$\mathbf{x} \sim p(\mathbf{x}|\mathbf{y})$

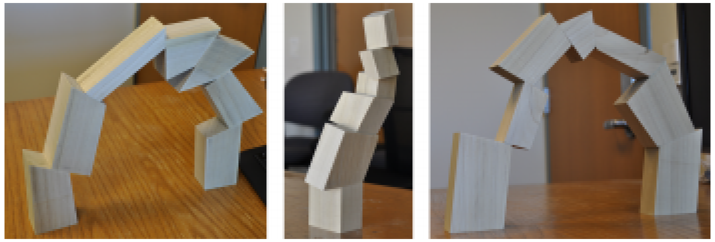
$\mathbf{x}$	$\mathbf{y}$
program source code	program output

Perov and Wood.
"Learning Probabilistic Programs."
arXiv:1407.2646 (2014).

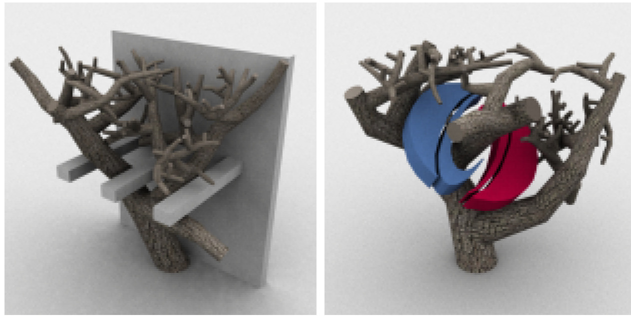
24

Application: Constrained Stochastic Simulation

Stable Static Structures



Procedural Graphics

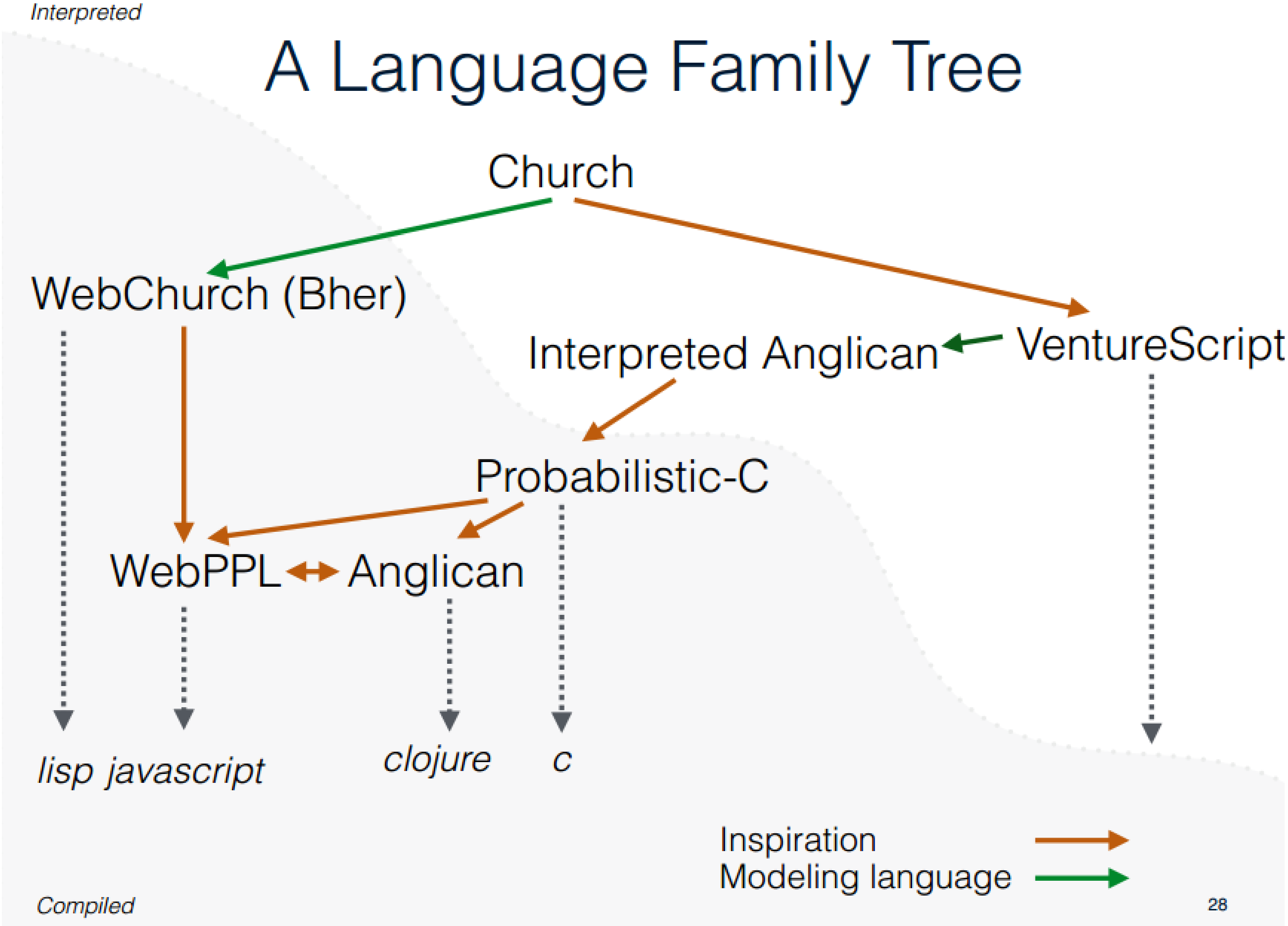


$\mathbf{x}$	$\mathbf{y}$
simulation	constraint

Ritchie, Lin, Goodman, & Hanrahan.
Generating Design Suggestions under Tight Constraints with Gradient-based Probabilistic Programming.
In Computer Graphics Forum, (2015)

Ritchie, Mildenhall, Goodman, & Hanrahan.
"Controlling Procedural Modeling Programs with Stochastically-Ordered Sequential Monte Carlo." 25
SIGGRAPH (2015)

A Language Family Tree



Conclusion

- Central question according to Josh Tenenbaum: How does the mind get so much from so little, in learning and reasoning about objects, categories, causes, scenes, events?
- Bayesian inference in probabilistic generative models.
- Probabilistic models defined over a range of structured representations: graphs, grammars, schemas, predicate logic…
- Hierarchical models, with inference at multiple levels of abstraction.
- Probabilistic programs: computationally universal representations for causal processes that are relational, recursive, composable.
- Towards a computational theory of human common sense.
- How can these theories be used to perceive, reason, predict, plan, learn and communicate …
- A lot to do at the intersection of cognitive science and machine learning …

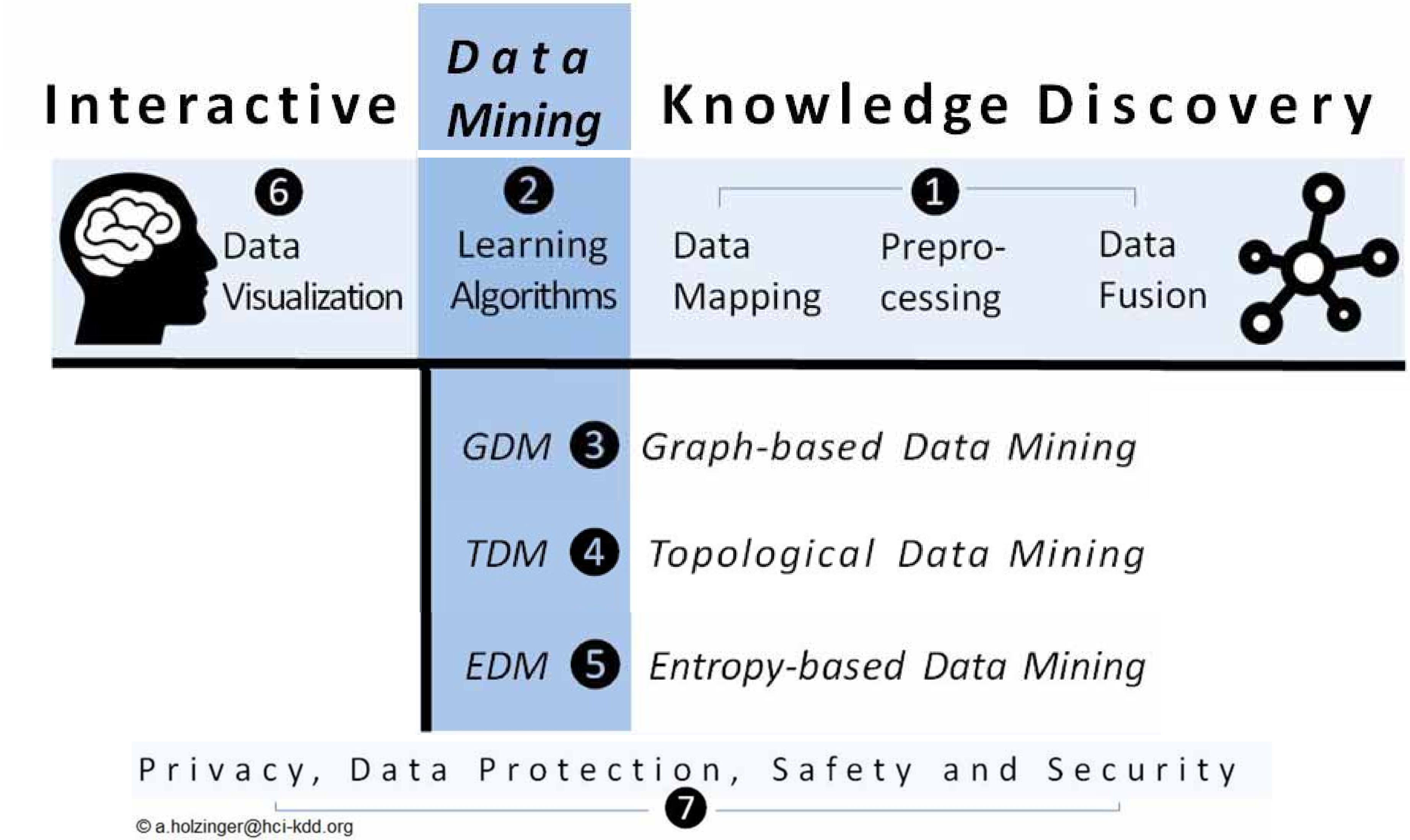
Conclusion

- **Probabilistic programming** is enabling to do things that would otherwise be impossible.
- **Inference > Probabilistic Programming > New Models > move forward ML > understand intelligence!**
- **(what did Demis Hassabis from Google Deep Mind say as their grand goal? ;-)**

concerted effort

international
without boundaries …

ML-Algorithms are key but needs also concerted effort



Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: Cognitive Science meets Machine Learning. IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

Thank you!

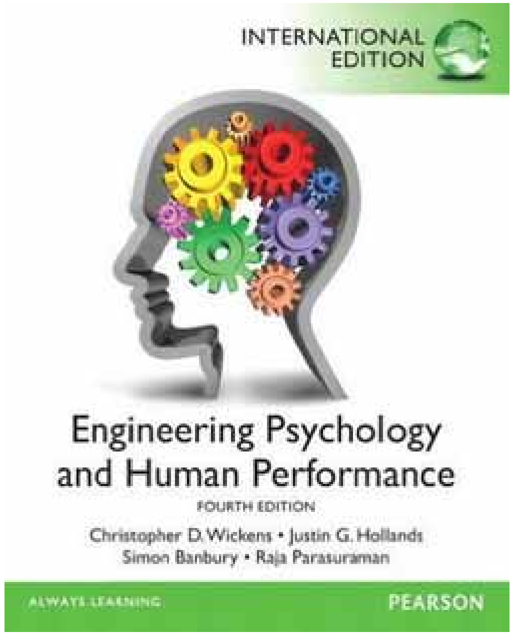
08 Questions

Sample Questions

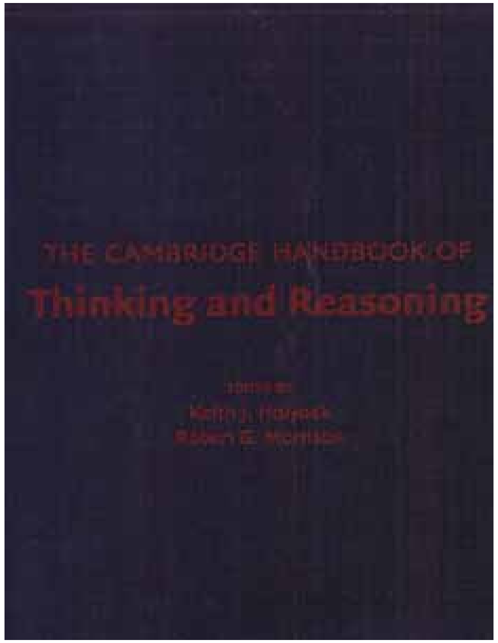
- Why is Cognitive Science important for Machine Learning?
- What is the mission statement of Google Deepmind?
- Describe the human information processing model of Atkinson & Shiffrin (1971)?
- Why is attention so central in cognition?
- How did we define knowledge?
- Why is decision making relevant for health informatics?
- What is probabilistic programming?
- What is reasoning?
- What can we do with probabilistic graphical models?
- What is the advantage of factor graphs?

09 Appendix

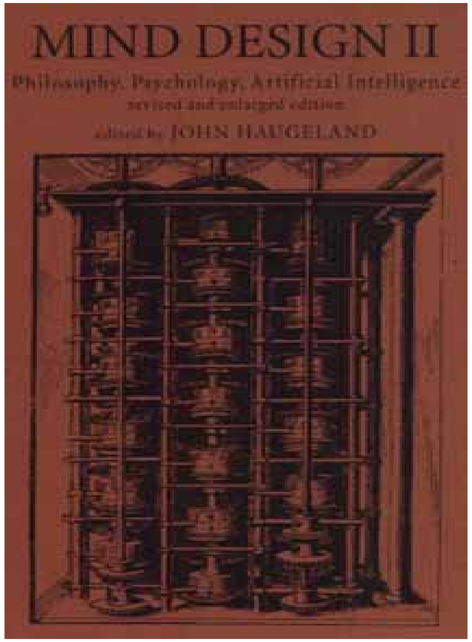
Recommended Books 1 (General Introductions)



Wickens, C. D., Hollands, J. G., Banbury, S. & Parasuraman, R. 2012. Engineering Psychology & Human Performance, 4th Edition, Boston et al., Pearson.



Holyoak, K. J. & Morrison, R. G. 2005. The Cambridge handbook of thinking and reasoning, Cambridge University Press.



Haugeland, J. 1997. Mind design II: philosophy, psychology, artificial intelligence, MIT press.

Related Conferences

- Year (2015)
- Help
- My Registrations
- Profile
- Sponsorship
- Exhibitor info
- Publications
- Future Meetings
- Women in Machine Learning

NIPS

NIPS 2015

Monday December 07 -- Saturday December 12, 2015
Palais des Congrès de Montréal, Montréal CANADA

Dates | Calls | Program Books | Schedule | NIPS Foundation | Awards

View NIPS 2016 » | View Earlier Meetings »

Follow | f | #nips2015

Online Videos

Videos are available for Tutorials, Invited Speakers, Workshops and Spotlights. Thanks to Microsoft for shooting, editing and hosting the videos. Check here for updates.

Tutorials

- *Introduction to Reinforcement Learning with Function Approximation*
- *High-Performance Hardware for Machine Learning*
- *Large-Scale Distributed Systems for Training Neural Networks*
- *Monte Carlo Inference Methods*
- *NIPS Deep Learning*
- *Probabilistic Programming*

Invited Speakers

- Zoubin Ghahramani, U. of Cambridge
- Mitsuo Kawato, ATR
- Asu Ozdaglar, MIT
- Haim Sompolinsky, Hebrew U.
- Robert Tibshirani, Stanford
- Vladimir Vapnik, Facebook AI Research, New York

Demonstrations

Symposia

Workshops

Related Journals, example (1)

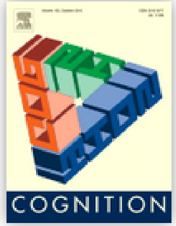
Cognition

International Journal of Cognitive Science

Editor-in-Chief: Steven Sloman

View Editorial Board

Supports Open Access



ISSN: 0010-0277

- Guide for Authors
- Submit Your Paper
- Track Your Paper
- Order Journal
- View Articles

Journal Metrics

Source Normalized Impact per Paper (SNIP): 1.676

SCImago Journal Rank (SJR): 2.770

Impact Factor: 3.411

5-Year Impact Factor: 4.308

Cognition is an international journal that publishes theoretical and experimental papers on the study of the mind. It covers a wide variety of subjects concerning all the different aspects of **cognition**, ranging from biological and experimental studies to formal analysis. Contributions from the fields of **psychology**, **neuroscience**, **linguistics**, **computer science**, **mathematics**, **ethology** and **philosophy** are welcome in this journal provided that they have some bearing on the functioning of the mind. In addition, the journal serves as a forum for discussion of social and political aspects of cognitive science.

Papers will be selected on the basis of their scientific quality and degree of innovation. A paper's theoretical relevance to cognition, overall soundness of the argument and degree of empirical motivation, especially from converging sources, are more important than adherence to specific methodological principles. Because *Cognition* enjoys a wide readership from many disciplines, authors should...

Read more

This journal supports the following content innovations

- AudioSlides
- Interactive Plot Viewer

http://www.journals.elsevier.com/cognition

Related Journals, example (2)

AMERICAN PSYCHOLOGICAL ASSOCIATION

About APA | Topics | Publications & Databases | Psychology Help Center | News & Events

Home // Publications & Databases // APA and Affiliated Journals // Psychological Review

Publications: Books | Children's Books | Databases | Journals | Magazines & Newsletters

Psychological Review®

Editor: Keith J. Holyoak, PhD

ISSN: 0033-295X

eISSN: 1939-1471

Published: six times, beginning in January

Impact Factor: 7.581

Psychology - Multidisciplinary: 5 of 129

View Table of Contents and Online First Publication

Read Sample Articles

Journal Snapshot

Advertising Information

5-Year Impact Factor: 10.872

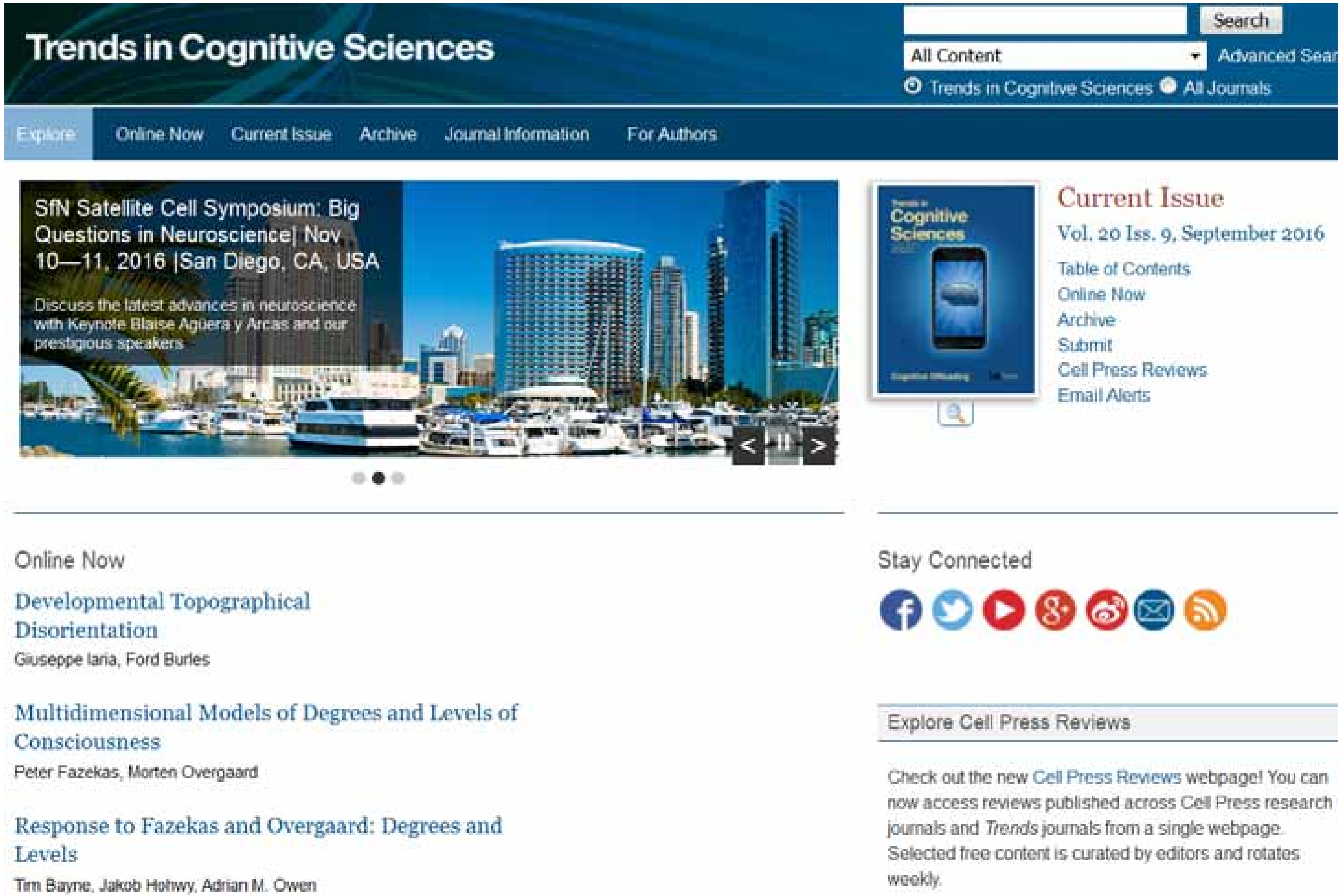
CABS 2015 Academic Journal Guide: Grade 4 (top-ranked)

- Editorial by Keith J. Holyoak, 2016 (PDF, 13KB)
- Read "Believing Superstitions That You Know Aren't True" from APA Journals Article Spotlight

http://www.apa.org/pubs/journals/rev/

Related Journals, example (3) (access from TUG, KFU)

Editorial Enquiries
Trends in Cognitive Sciences
Cell Press
50 Hampshire St.
5th Floor
Cambridge, MA 02139
USA
Tel: +1 617 397 2817
Fax: +1 617 397 2810
E-mail: tics@cell.com



http://www.cell.com/trends/cognitive-sciences/home

Solutions to the Quiz Questions

1. People are awesome – this cannot be done by any machine learning device – no robot to date can do that complex behaviour
2. No, on Mars computers are better to date
3. No, driverless cars can drive when humans are not able to drive
4. No, chess computers are playing better, but sometimes you can win through illogical behaviour (Spock)
5. Partly, vacuum cleaning is not a sophisticated task
6. NLP partly, primitive tasks can be done by computers
7. Humanoid robotics – also far from the reality, as for example in Ex Machina – e.g in Fukusima
8. Image understanding is hard for a machine if there are not thousands of samples before – humans can get out so much from so little
9. In Rn no human has a chance
10. Mathematics partly – creativity is better in humans
11. Pattern recognition – partly humans are good but in big data machines are much better
12. In NP-hard problems humans have a chance via heuristics – humans have creativity! See foldit

Additional Reading

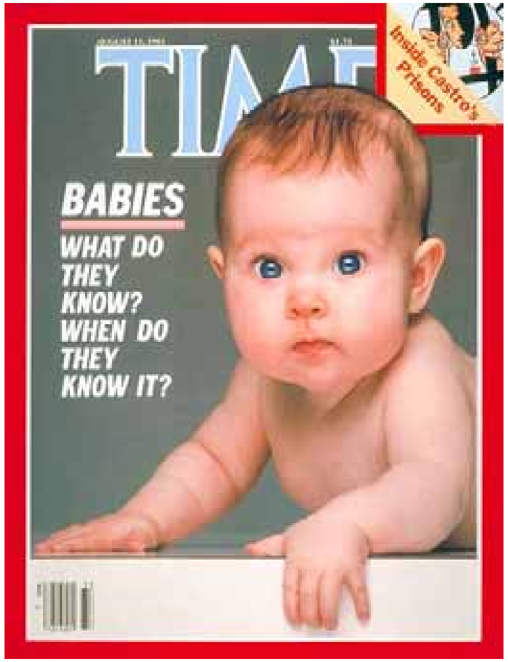
- **History of Probability Theory**
- Franklin, J. *The Science of Conjecture: Evidence and Probability Before Pascal.* John Hopkins University Press, 2001.
- Jaynes, E. T. *Probability Theory: The Logic of Science.* Cambridge University Press, 2003.
- **Probabilistic Reasoning**
- Gigerenzer, G., and D. J. Murray. *Cognition as Intuitive Statistics*. Hillsdale, NJ: Erlbaum, 1987.
- Gilovich, T., D. Griffin, and D. Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, 2002.
- Kahneman, D., P. Slovic, and A. Tversky, eds. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.
- **Bayesian Networks**
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman, San Mateo, CA, 1988.
- Breese, J. S. "Construction of Belief and Decision Networks." *Computational Intelligence* 8, 4 (1992): 624–647.
- F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. "Statistical Foundations for Default Reasoning." *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI).* Chambery, France, August 1993, pp. 563-569.
- **Multiple-Instance Bayesian Networks**
- Pasula, H., and S. Russell. "Approximate Inference for First-order Probabilistic Languages." *IJCAI-0*1. Seattle, WA, 2001, pp. 741–748.
- Halpern, J. Y. "An Analysis of First-order Logics of Probability." *Artificial Intelligence* 46, 3 (1990): 311–350.
- D. Koller, and A. Pfeffer. "Object-Oriented Bayesian Networks." *Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)*. Providence, Rhode Island, 1997, pp. 302-313.

- **Causal Reasoning**
- Waldmann, M. R. "Competition among Causes but not Effects in Predictive and Diagnostic Learning." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (2000): 53-76.
- Ahn, W., and M. Dennis. "Induction of Causal Chain." *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society.* Lawrence Erlbaum Associates, NJ: Mahwah, 2000.
- Dennis, M. J., and W. Ahn. "Primacy in Causal Strength Judgments." *Memory & Cognition* 29 (2001): 152-164
- Cheng, P. W. "From Covariation to Causation: A Causal Power Theory." *Psychological Review* 104 (1997): 367-405.
- Novick, L. R., and P. W. Cheng. "Assessing Interactive Causal Influence." *Psychological Review.* (in press)
- **Causal Models**
- Pearl, J. *Causality: Models, Reasoning, and Inference.* New York: Cambridge University Press, 2000.
- Glymour, C. "Learning, Prediction and Causal Bayes Nets." *Trends in Cognitive Science 7* (2003): 43-48.
- Zhang, N., and D. Poole. "Exploiting Causal Independence in Bayesian Network Inference." *Journal of Artificial Intelligence Research* 5 (1996): 301–328.

Additional Reading (2)

- **Causal Theories and Mechanisms**
- Waldmann, M. R. "Knowledge-based Causal Induction." *The Psychology of Learning and Motivation, Vol. 34: Causal Learning.* Edited by D. R. Shanks, K. J. Holyoak, and D. L. Medin. San Diego: Academic Press, 1996, pp. 47-88.
- Ahn, W., C. W. Kalish, D. L. Medin, and S. A. Gelman. "The Role of Covariation vs. Mechanism Information in Causal Attribution." *Cognition* 54 (1995): 299-352.
- Ahn, W., L. Novick, and N. S. Kim. "Understanding it Makes it More Normal: Causal Explanations Influence Person Perception." *Psychonomic Bulletin and Review*. (in press)
- **Rational Analysis**
- Anderson, J. R. *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1990.
- Ahn, W., and L. M. Graham. "The Impact of Necessity and Sufficiency on Information Choices in the Wason Four-card Selection Task." *Psychological Science* 10 (1999): 237-242.
- Oaksford, M., and N. Chater. "A Rational Analysis of the Selection Task as Optimal Data Selection." *Psychological Review* 101 (1994): 608-631.
- Oaksford M., and N. Chater, eds. *Rational Models of Cognition*. Oxford University Press, 1998.
- Sperber, D., F. Cara, and V. Girotto. "Relevance Theory Explains the Selection Task". *Cognition* 57 (1995): 31-95.

Spare Time Reading



http://content.time.com/time/magazine/article/0,9171,1580382,00.ht
ml

http://content.time.com/time/magazine/article/0,9171,30198,00.html