Computers and Thought

Guy Van den Broeck

IJCAI

August 16, 2019

Outline

1. What would 2011 junior PhD student Guy think?

   …*please help me make sense of this field*…

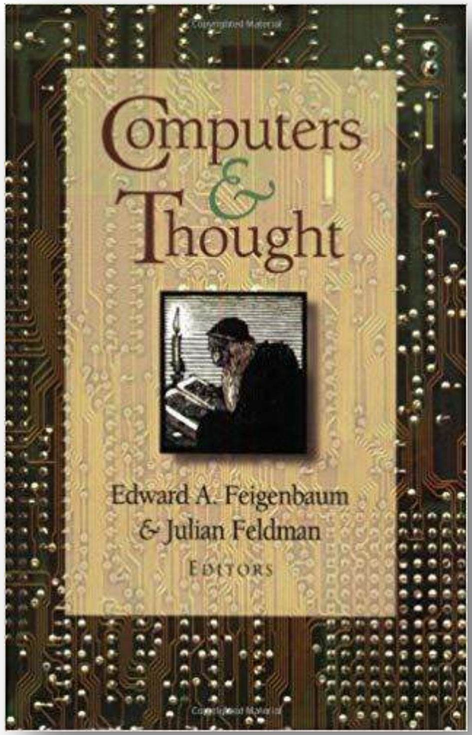
2. What do I work on and why?
   – High-level probabilistic reasoning
   – A new synthesis of learning and reasoning

3. Personal thank you messages

The AI Dilemma of 2019

Deep learning
approaches the problem of designing intelligent machines by postulating a large number of very simple information processing elements, arranged in a [.] network, and certain processes for facilitating or inhibiting their activity.

Knowledge representation and reasoning
take a much more macroscopic approach [.]. They believe that intelligent performance by a machine is an end difficult enough to achieve without “starting from scratch” , and so they build into their systems as much complexity of information processing as they are able to understand and communicate to a computer.

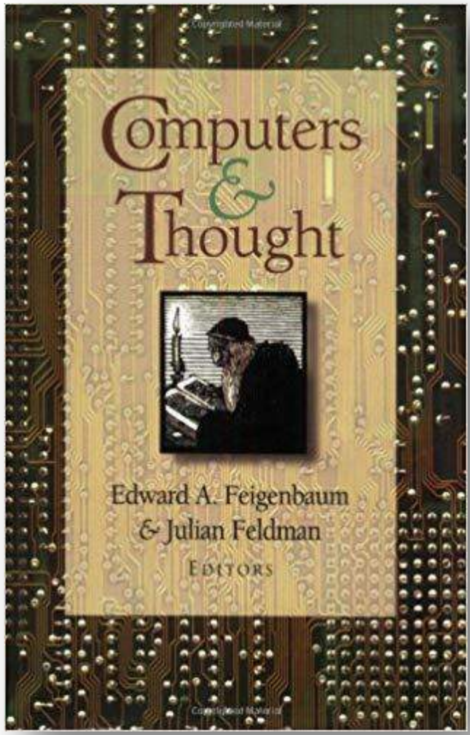


Edward Feigenbaum and Julian Feldman

The AI Dilemma of ~~2019~~ **1963**

Neural cybernetics
approaches the problem of designing intelligent machines by postulating a large number of very simple information processing elements, arranged in a [.] network, and certain processes for facilitating or inhibiting their activity.

Cognitive model builders
take a much more macroscopic approach [.]. They believe that intelligent performance by a machine is an end difficult enough to achieve without "starting from scratch", and so they build into their systems as much complexity of information processing as they are able to understand and communicate to a computer.

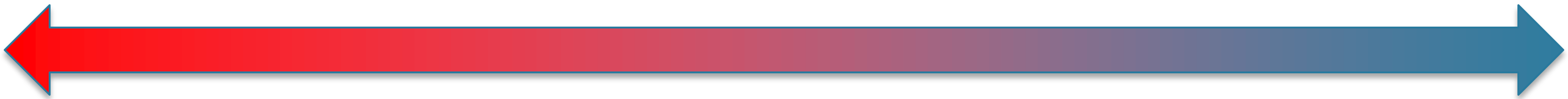


Edward Feigenbaum and Julian Feldman

The AI Dilemma

Pure Logic **Pure Learning**

The AI Dilemma



Pure Logic **Pure Learning**

- Slow thinking: deliberative, cognitive, model-based, extrapolation
- Amazing achievements until this day

The AI Dilemma

Pure Logic **Pure Learning**

- Slow thinking: deliberative, cognitive, model-based, extrapolation
- Amazing achievements until this day
- "*Pure logic is brittle*"
  noise, uncertainty, incomplete knowledge, …

The AI Dilemma

Pure Logic **Pure Learning**

- Fast thinking: instinctive, perceptive, model-free, interpolation
- Amazing achievements recently

The AI Dilemma

Pure Logic **Pure Learning**

- Fast thinking: instinctive, perceptive, model-free, interpolation
- Amazing achievements recently
- "*Pure learning is brittle*"

bias, algorithmic fairness, interpretability, explainability, adversarial attacks, unknown unknowns, calibration, verification, missing features, missing labels, data efficiency, shift in distribution, general robustness and safety

fails to incorporate a sensible model of the world

Knowledge vs. Data

- Where did the world knowledge go?
  - Python scripts
    - Decode/encode cleverly
    - Fix inconsistent beliefs
  - Rule-based decision systems
  - Dataset design
  - “a big hack” (with author’s permission)
- In some sense we went backwards

  Less principled, scientific, and intellectually satisfying ways of incorporating knowledge

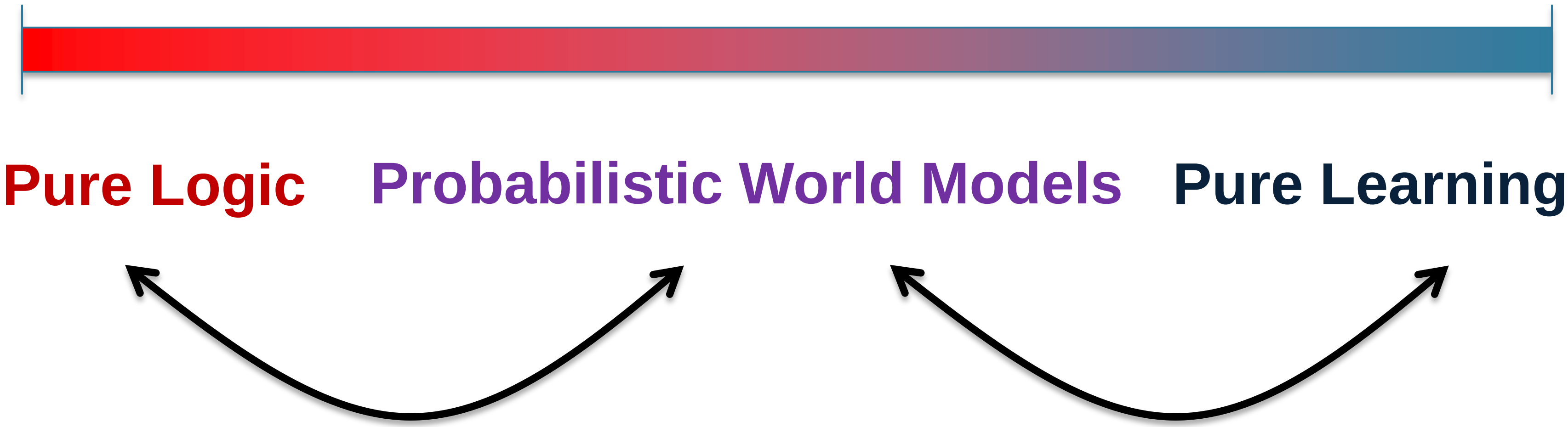
The **FALSE** AI Dilemma

So all hope is lost?

Probabilistic World Models

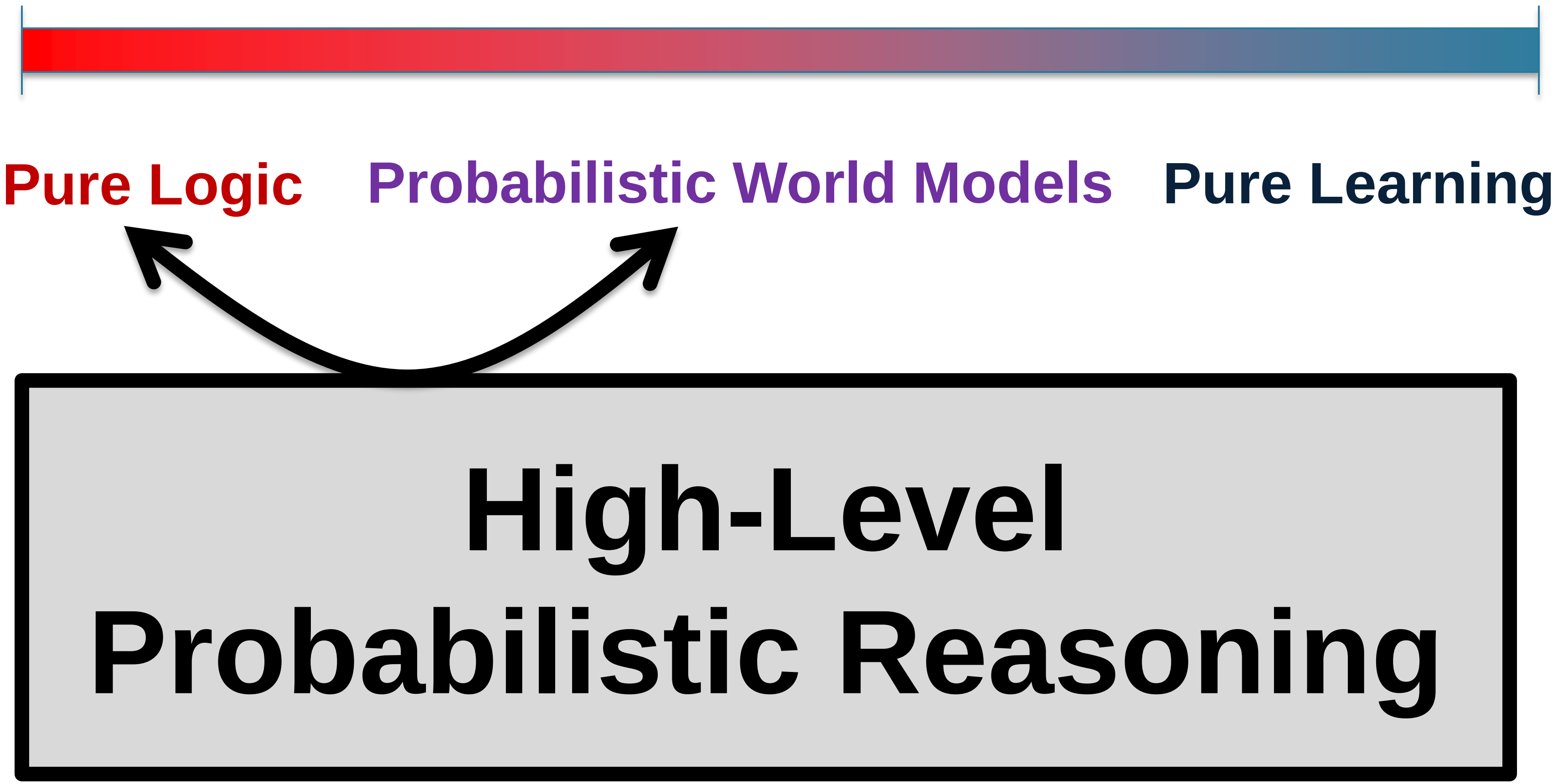
- Joint distribution P(X)
- Wealth of representations:
  can be causal, relational, etc.
- Knowledge + data
- Reasoning + learning

Then why isn't everything solved?

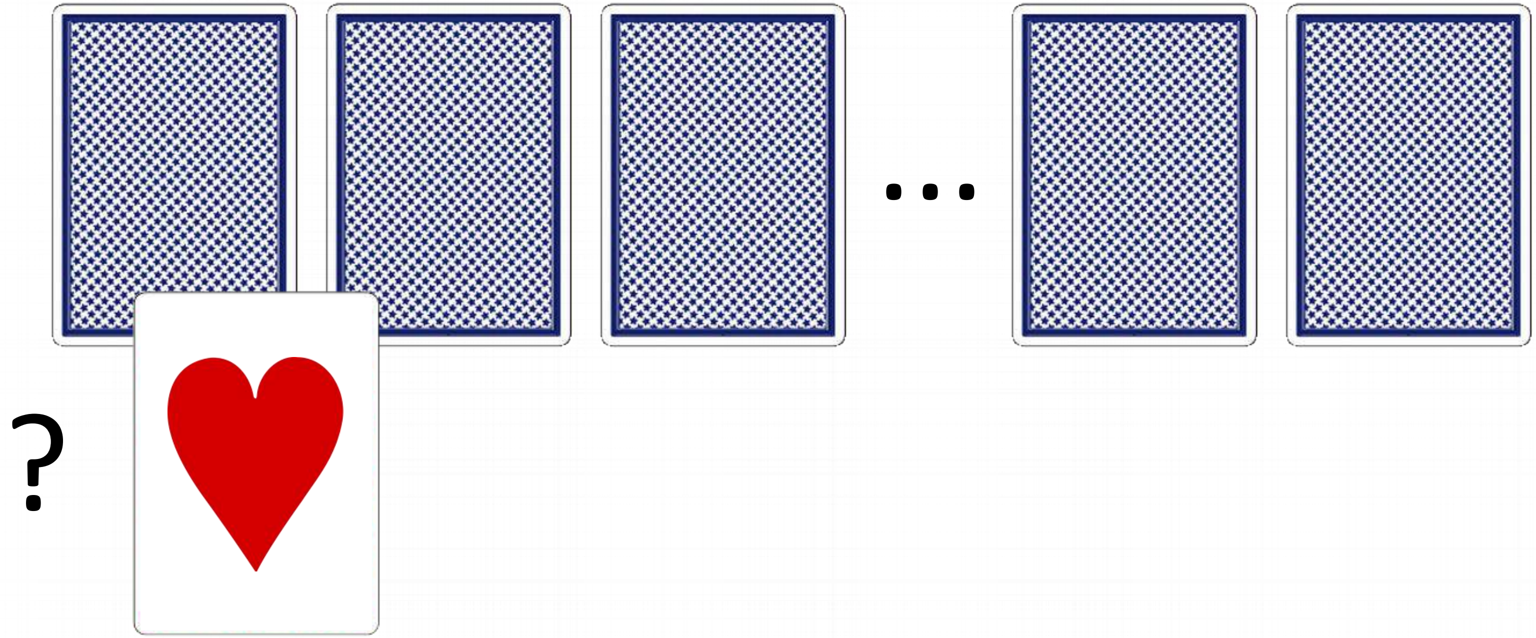


What did we gain?

What did we lose along the way?


Pure Logic
Probabilistic World Models
Pure Learning
High-Level
Probabilistic Reasoning

Simple Reasoning Problem

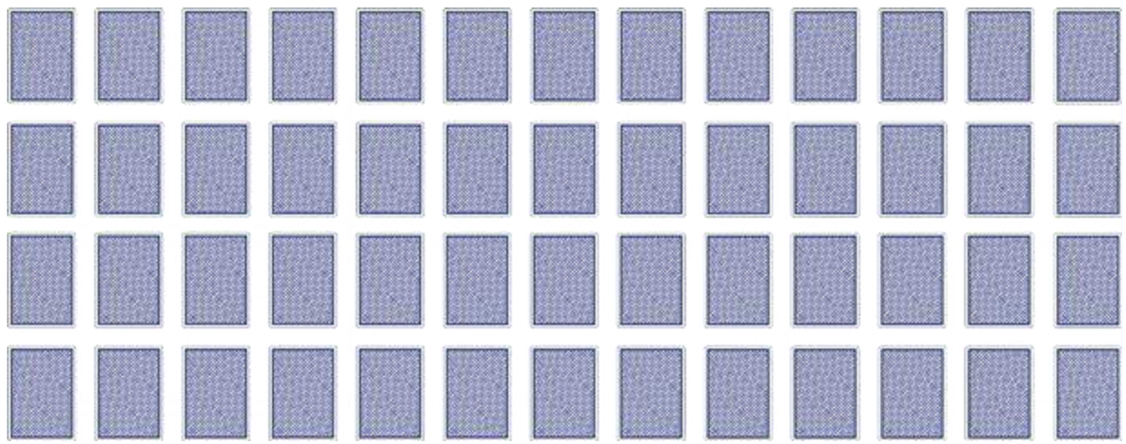


Probability that first card is Hearts? 1/4

Automated Reasoning

Let us automate this:

1. Probabilistic graphical model (e.g., factor graph)

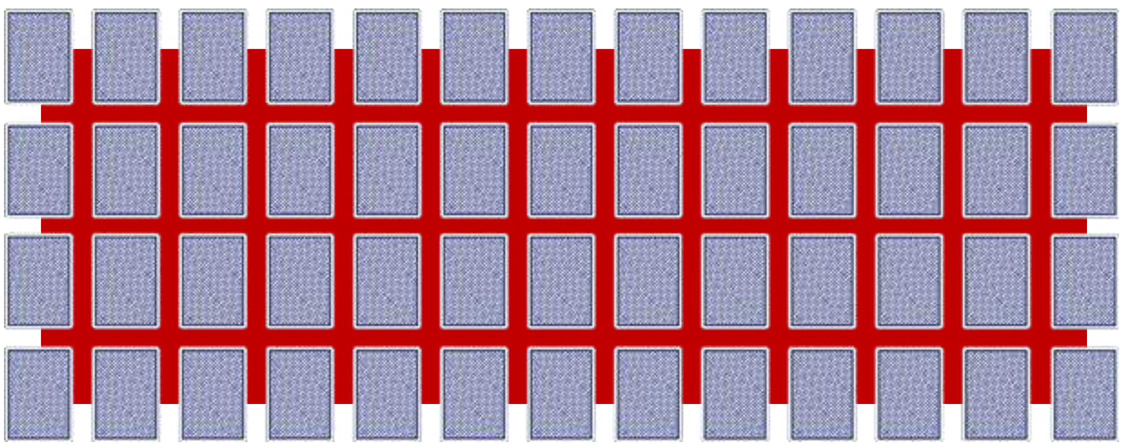


2. Probabilistic inference algorithm
   (e.g., variable elimination or junction tree)

Automated Reasoning

Let us automate this:

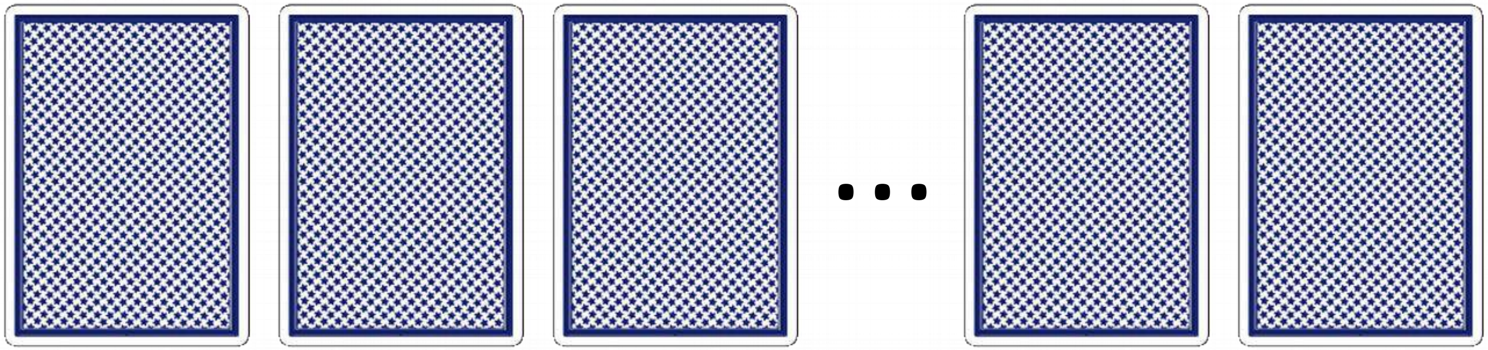
1. Probabilistic graphical model (e.g., factor graph) is fully connected!



(artist's impression)

2. Probabilistic inference algorithm (e.g., variable elimination or junction tree) builds a table with **52^{52}** rows

Tractable High-Level Reasoning

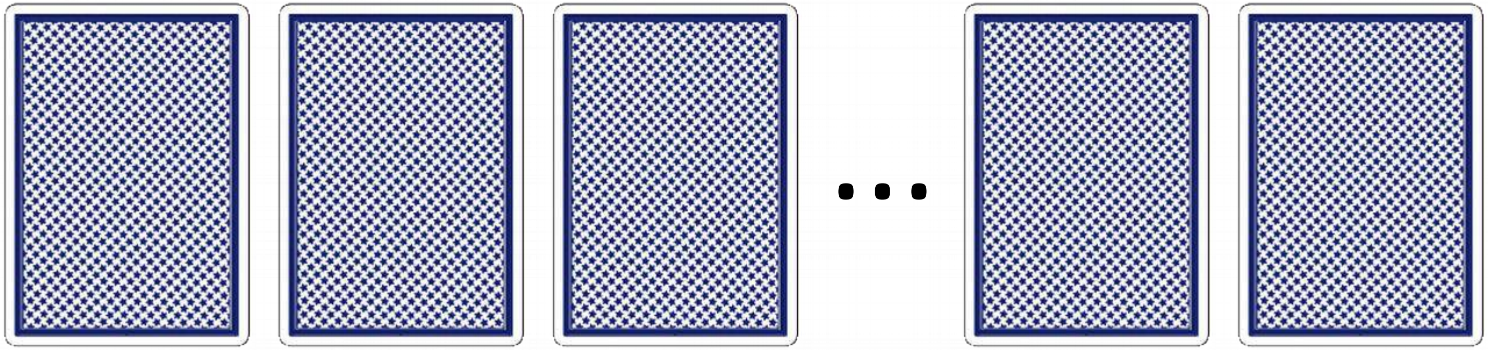


What's going on here?

Which property makes reasoning tractable?

- High-level (first-order) reasoning
- Symmetry
- Exchangeability

⇒ Lifted Inference

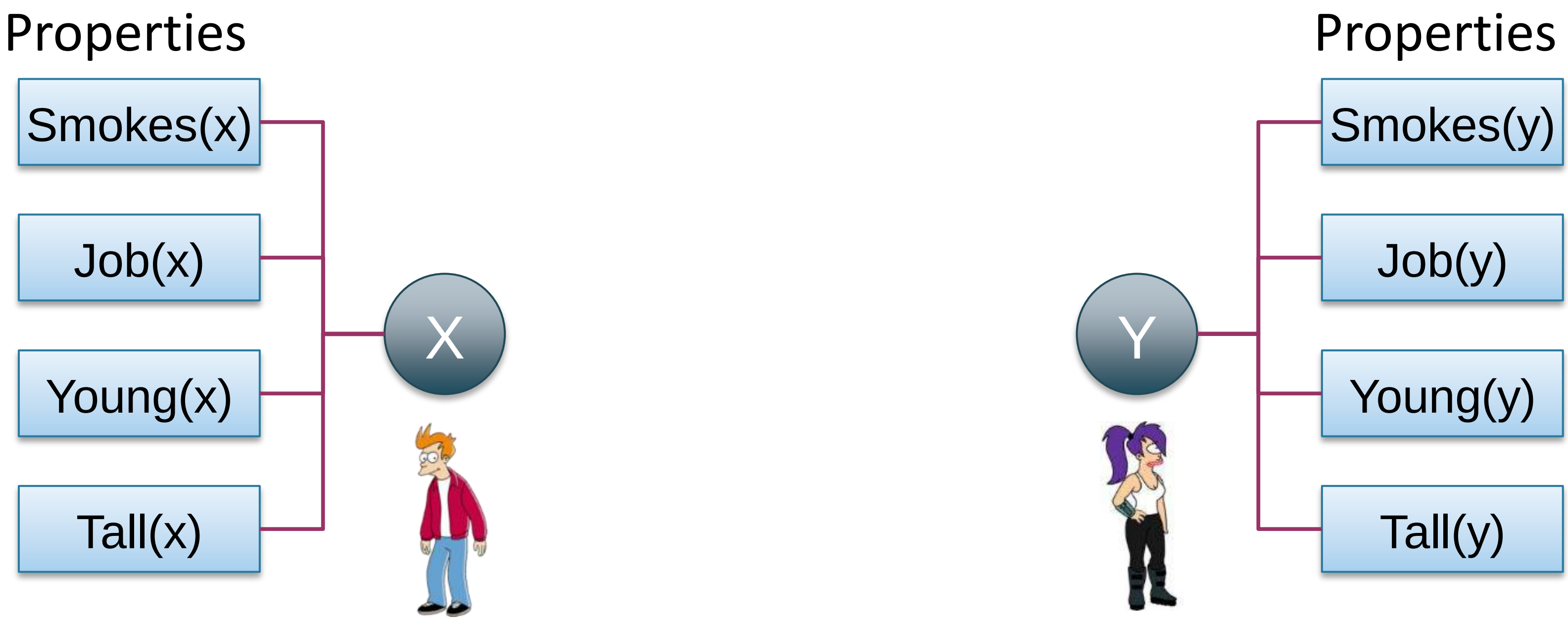


Model distribution at first-order level:

$$\forall p, \exists c, \mathrm{Card}(p,c)$$
$$\forall c, \exists p, \mathrm{Card}(p,c)$$
$$\forall p, \forall c, \forall c', \mathrm{Card}(p,c) \wedge \mathrm{Card}(p,c') \Rightarrow c = c'$$

Can we now be efficient
in the size of our domain?

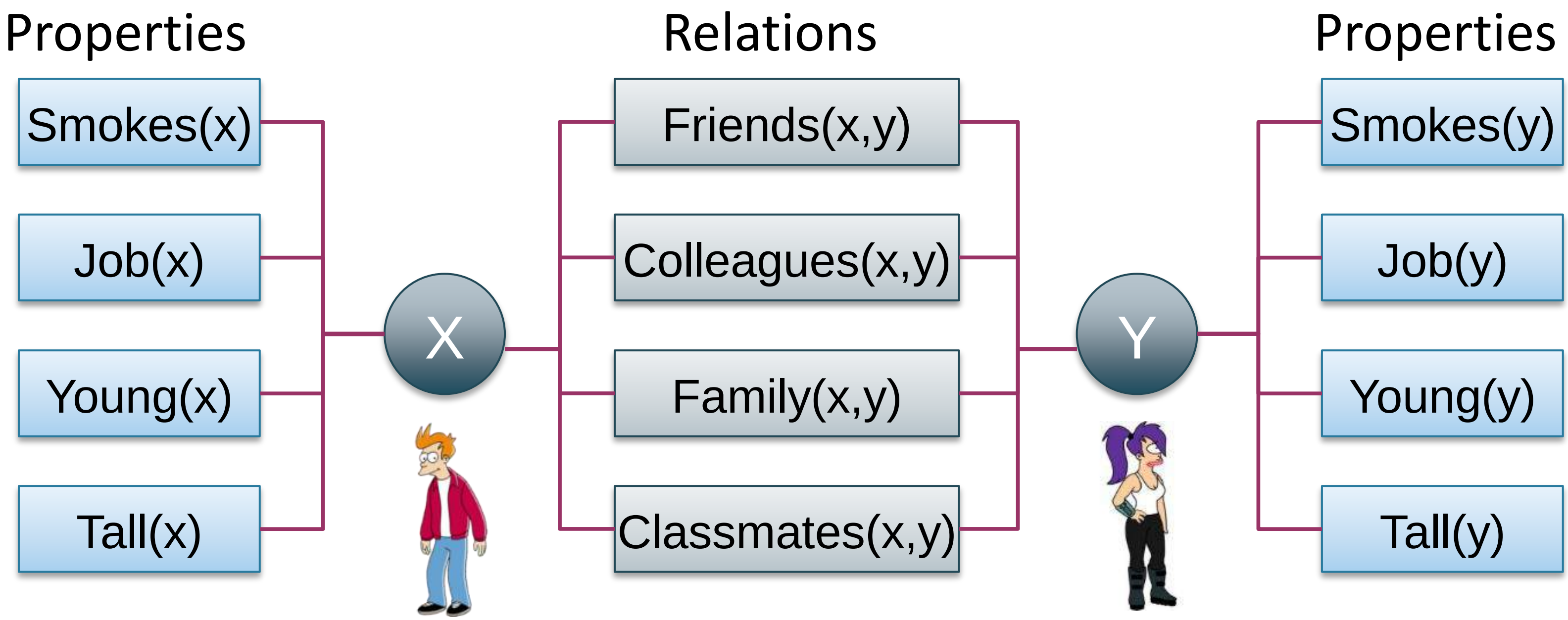
How does this relate to learning?



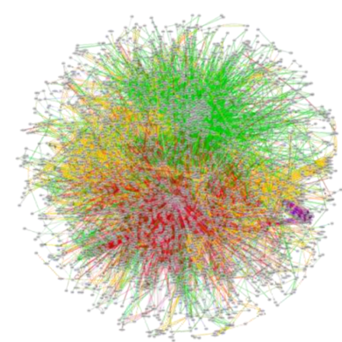
i.i.d. assumption

independent and identically distributed

Relational Learning



"Smokers are more likely to be friends with other smokers."
"Colleagues of the same age are more likely to be friends."
"People are either family or friends, but never both."
"If X is family of Y, then Y is also family of X."
"Universities in California are more likely to be rivals."



Lifted Inference Example: Counting Possible Worlds

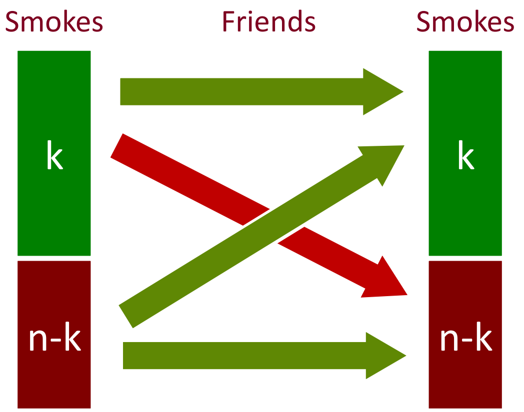
∀x ,y ∈ People: Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

- If we know **D** precisely: who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...

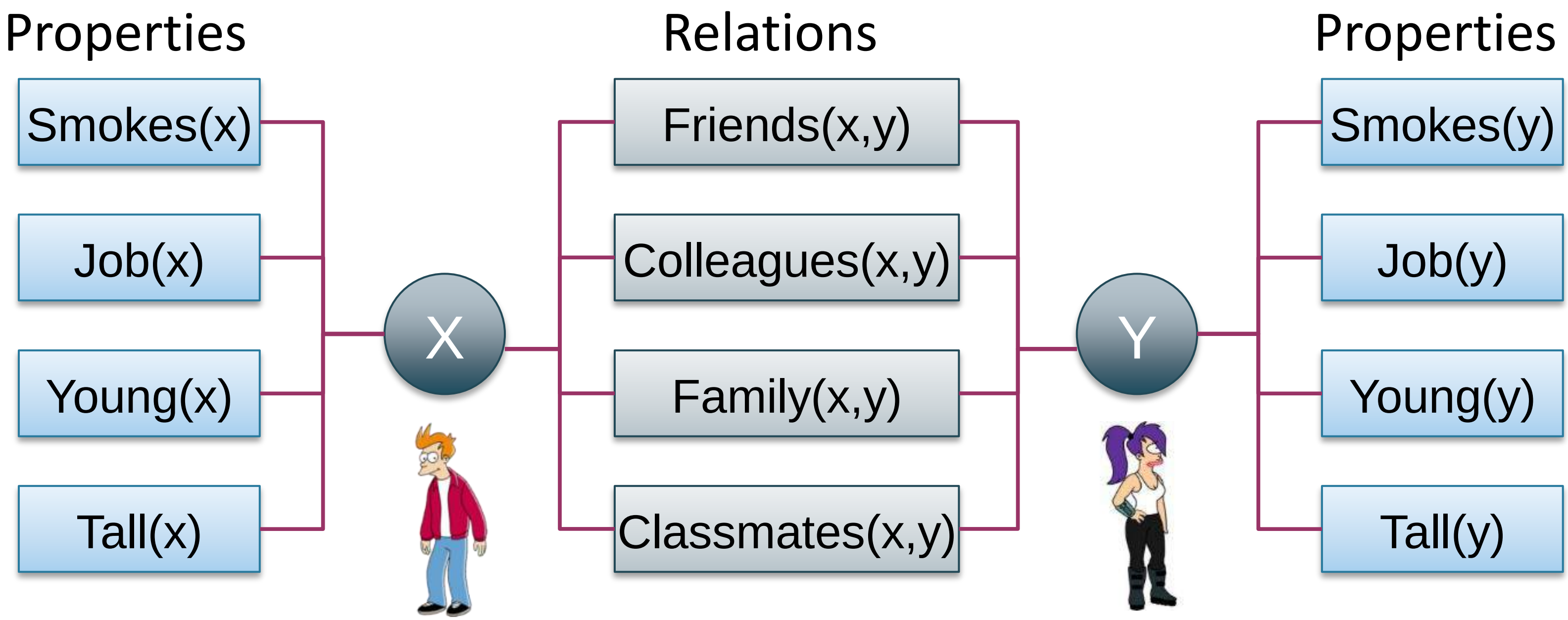
→ $2^{n^2-k(n-k)}$ worlds



- If we know that there are *k* smokers? → $\binom{n}{k} 2^{n^2-k(n-k)}$ worlds

- In total… → $\sum_{k=0}^{n} \binom{n}{k} 2^{n^2-k(n-k)}$ worlds

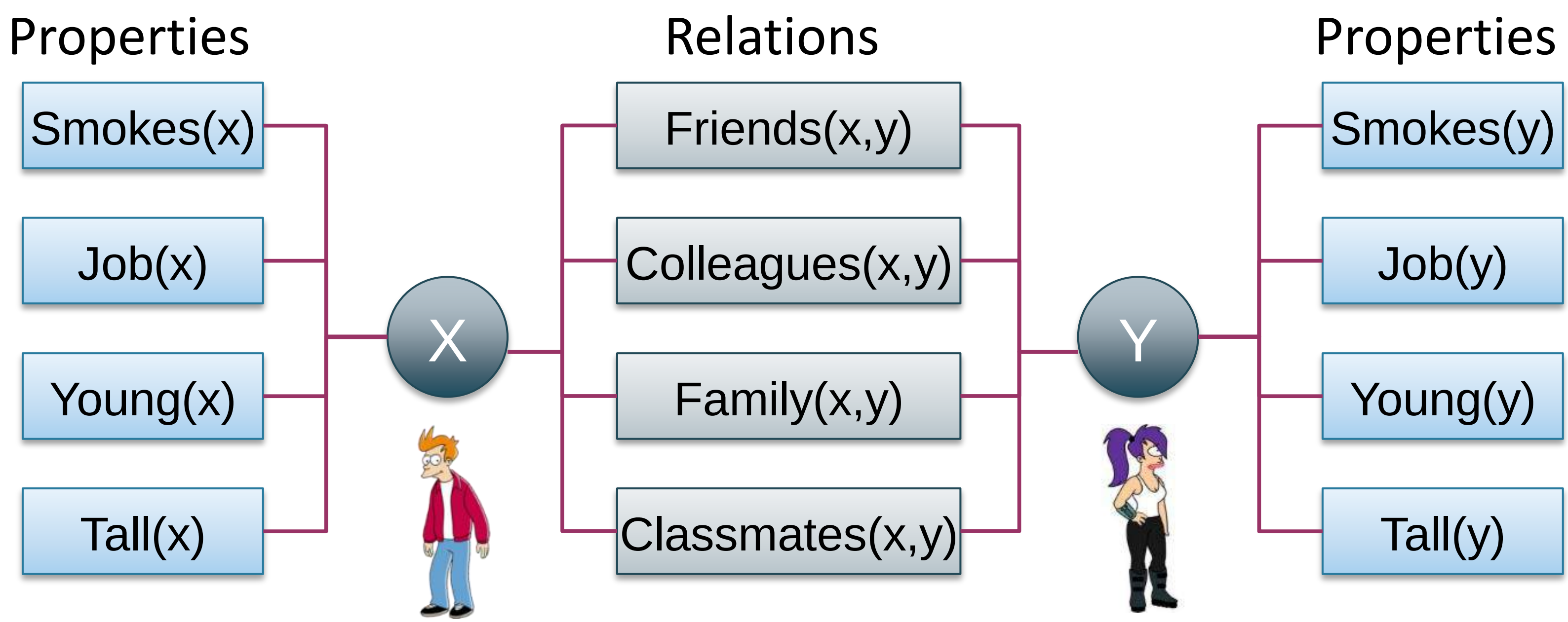
FO^2 is Liftable!



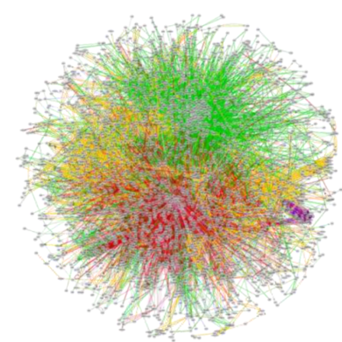
Theorem: Model counting for FO^2 in polynomial time in the number of constants/nodes/entities/people/cards.

Corollary: Partition functions efficient to compute in 2-variable Markov logic, relational factor graphs, etc.

FO^2 is Liftable!



"Smokers are more likely to be friends with other smokers."
"Colleagues of the same age are more likely to be friends."
"People are either family or friends, but never both."
"If X is family of Y, then Y is also family of X."
"Universities in California are more likely to be rivals."



Can Everything Be Lifted?

Theorem: There exists an FO^3 model Θ_1 for which just counting possible worlds is $\#P_1$-complete in the domain size.

What about learning?

- Learn better models faster
- Tractability is a great inductive bias!

	IMDb			*UWCSE*		
	Baseline	Lifted Weight Learning	Lifted Structure Learning	Baseline	Lifted Weight Learning	Lifted Structure Learning
Fold 1	-548	-378	**-306**	-1,860	-1,524	**-1,477**
Fold 2	-689	-390	**-309**	-594	-535	**-511**
Fold 3	-1,157	-851	**-733**	-1,462	-1,245	**-1,167**
Fold 4	-415	-285	**-224**	-2,820	-2,510	**-2,442**
Fold 5	-413	-267	**-216**	-2,763	-2,357	**-2,227**

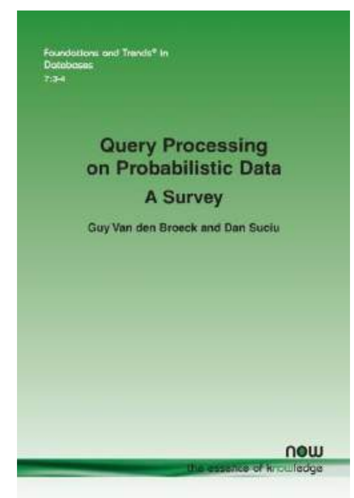
Pure Logic **Probabilistic World Models** **Pure Learning**

"A confluence of ideas,
a meeting place of two streams of thought"

Probabilistic Logic Programming
Prolog meets probabilistic AI

Probabilistic Databases
Databases meets probabilistic AI

Weighted Model Integration
SAT modulo theories meets probabilistic AI

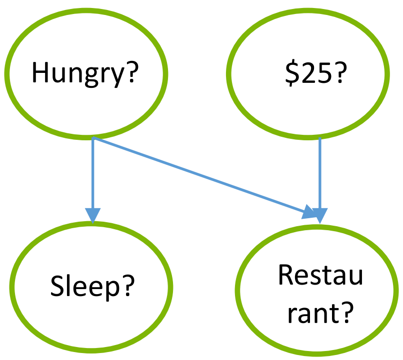


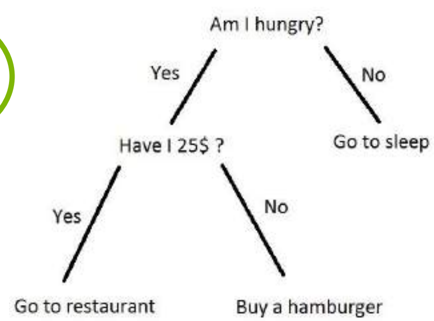
Pure Logic Probabilistic World Models Pure Learning

A New Synthesis of Learning and Reasoning

Another False Dilemma?

Classical AI Methods

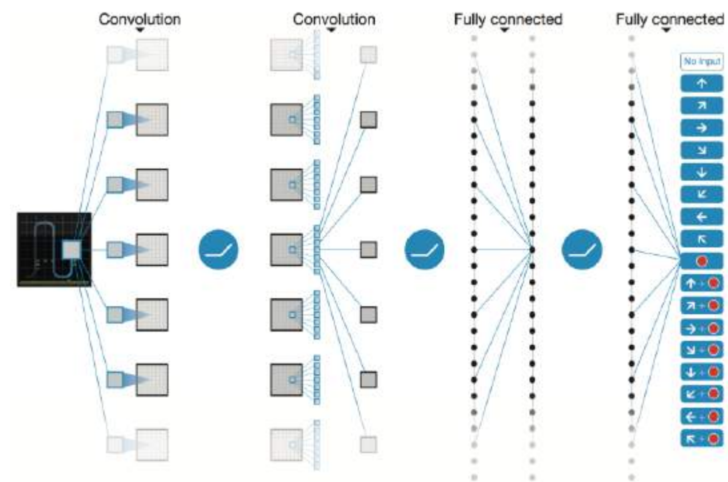




...

Clear Modeling Assumption
Well-understood

Neural Networks



“Black Box”
Empirical performance

Probabilistic Circuits

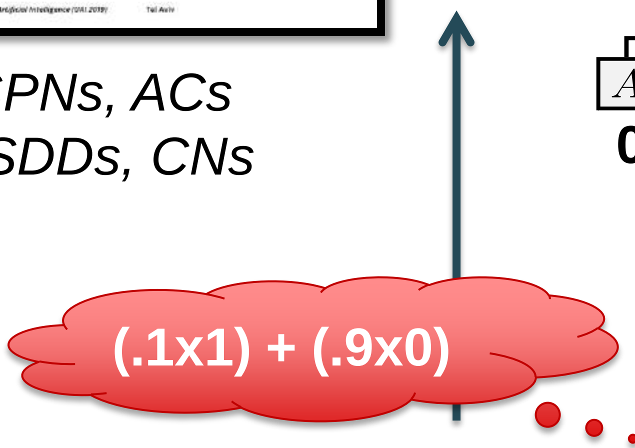

Tractable Probabilistic Models

Representations
Inference
Learning
Applications

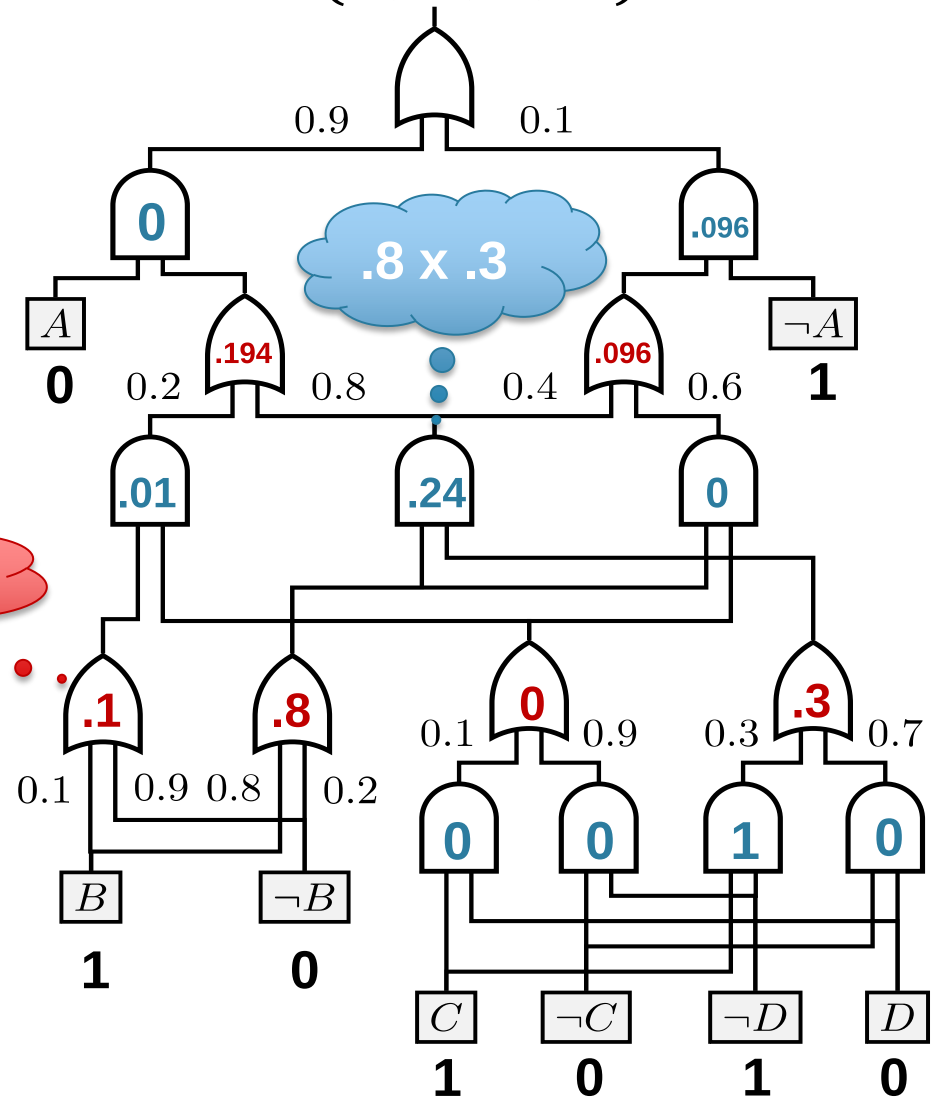
Antonio Vergari
University of California, Los Angeles
Nicola Di Mauro
University of Bari
Guy Van den Broeck
University of California, Los Angeles

July 22, 2019 - Conference on Uncertainty in Artificial Intelligence (UAI 2019) Tel Aviv

SPNs, ACs
PSDDs, CNs



$$\mathbf{Pr}(A, B, C, D) = \mathbf{0.096}$$



Input:

A	B	C	D
0	1	1	0

Properties, Properties, Properties!

- Read conditional independencies from structure
- Interpretable parameters (XAI)
  (conditional probabilities of logical sentences)
- Closed-form parameter learning
- Efficient reasoning (linear ☺)
  - Computing **conditional probabilities** Pr(x|y)
  - **MAP inference**: most-likely assignment to x given y
  - Even much harder tasks: expectations, KLD, entropy, logical queries, decision making queries, etc.

Probabilistic Circuits: Performance

Density estimation benchmarks: tractable vs. intractable

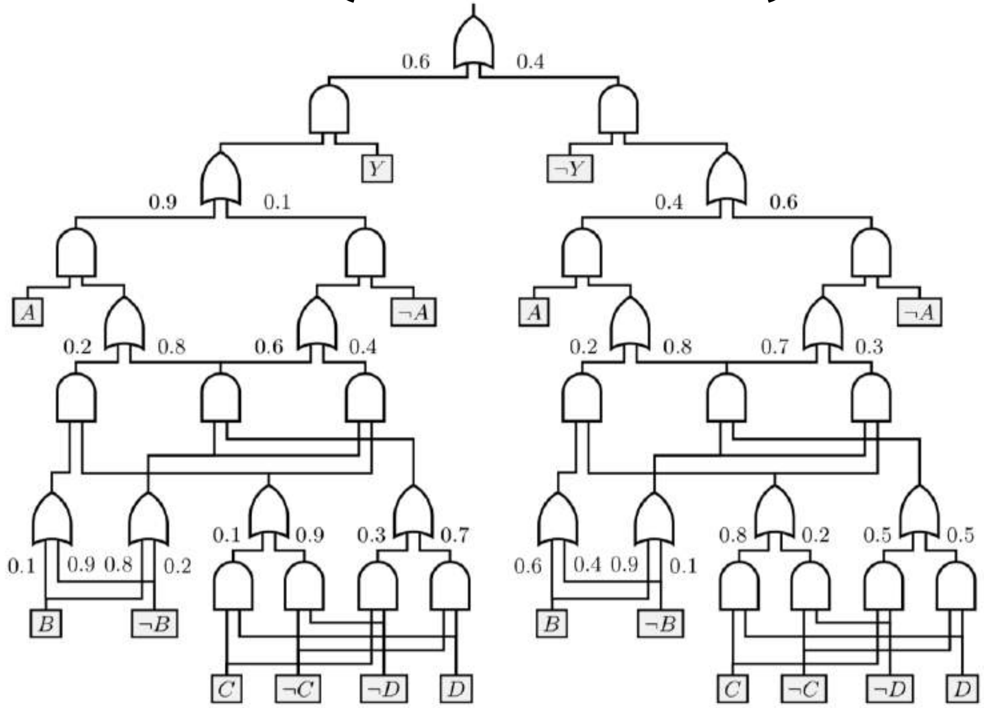
Dataset	best circuit	BN	MADE	VAE
nltcs	**-5.99**	-6.02	-6.04	**-5.99**
msnbc	**-6.04**	**-6.04**	-6.06	-6.09
kdd2000	-2.12	-2.19	**-2.07**	-2.12
plants	**-11.84**	-12.65	12.32	-12.34
audio	-39.39	-40.50	-38.95	**-38.67**
jester	-51.29	**-51.07**	-52.23	-51.54
netflix	-55.71	-57.02	-55.16	**-54.73**
accidents	-26.89	**-26.32**	-26.42	-29.11
retail	**-10.72**	-10.87	-10.81	-10.83
pumbs*	-22.15	**-21.72**	-22.3	-25.16
dna	**-79.88**	-80.65	-82.77	-94.56
Kosarek	**-10.52**	-10.83	-	-10.64
Msweb	-9.62	-9.70	**-9.59**	-9.73

Dataset	best circuit	BN	MADE	VAE
Book	-33.82	-36.41	-33.95	**-33.19**
movie	-50.34	-54.37	-48.7	**-47.43**
webkb	-149.20	-157.43	-149.59	**-146.9**
cr52	-81.87	-87.56	-82.80	**-81.33**
c20ng	-151.02	-158.95	-153.18	**-146.90**
bbc	**-229.21**	-257.86	-242.40	-240.94
ad	-14.00	-18.35	**-13.65**	-18.81

But what if I only want to classify?

$$\Pr(Y|A,B,C,D)$$

$$\cancel{\Pr(Y,A,B,C,D)}$$



Logistic Circuits

Comparable Accuracy with Neural Nets

Accuracy % on Dataset	Mnist	Fashion
Baseline: Logistic Regression	85.3	79.3
Baseline: Kernel Logistic Regression	97.7	88.3
Random Forest	97.3	81.6
3-layer MLP	97.5	84.8
RAT-SPN (Peharz et al. 2018)	98.1	89.5
SVM with RBF Kernel	98.5	87.8
5-Layer MLP	99.3	89.8
Logistic Circuit (binary)	97.4	87.6
Logistic Circuit (real-valued)	99.4	91.3
CNN with 3 conv layers	99.1	90.7
Resnet (He et al. 2016)	99.5	93.6

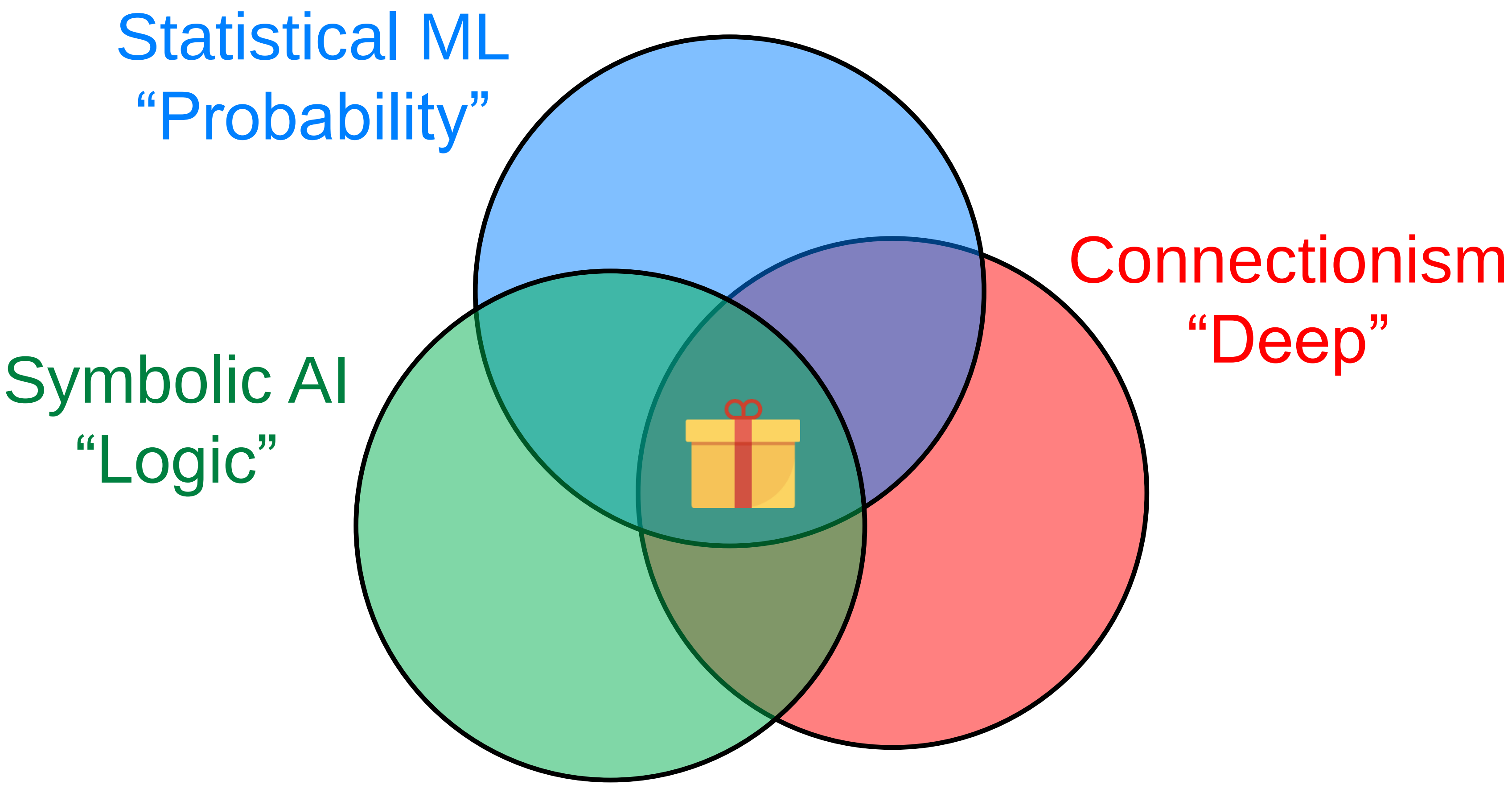
Significantly Smaller in Size

Number of Parameters	Mnist	Fashion
Baseline: Logistic Regression	<1K	<1K
Baseline: Kernel Logistic Regression	1,521 K	3,930K
Logistic Circuit (real-valued)	182K	467K
Logistic Circuit (binary)	268K	614K
3-layer MLP	1,411K	1,411K
RAT-SPN (Peharz et al. 2018)	8,500K	650K
CNN with 3 conv layers	2,196K	2,196K
5-Layer MLP	2,411K	2,411K
Resnet (He et al. 2016)	4,838K	4,838K

Better Data Efficiency

Accuracy % with % of Training Data	MNIST			Fashion		
	100%	10%	2%	100%	10%	2%
5-layer MLP	99.3	**98.2**	94.3	89.8	86.5	80.9
CNN with 3 Conv Layers	99.1	98.1	95.3	90.7	87.6	83.8
Logistic Circuit (Binary)	97.4	96.9	94.1	87.6	86.7	83.2
Logistic Circuit (Real-Valued)	**99.4**	97.6	**96.1**	**91.3**	**87.8**	**86.0**

Probabilistic & Logistic Circuits



Reasoning about
World Model + Classifier

"*Pure learning is brittle*"

bias, algorithmic fairness, interpretability, explainability, adversarial attacks, unknown unknowns, calibration, verification, missing features, missing labels, data efficiency, shift in distribution, general robustness and safety

fails to incorporate a sensible model of the world

- Given a learned predictor F(x)
- Given a probabilistic world model P(x)
- How does the world act on learned predictors?

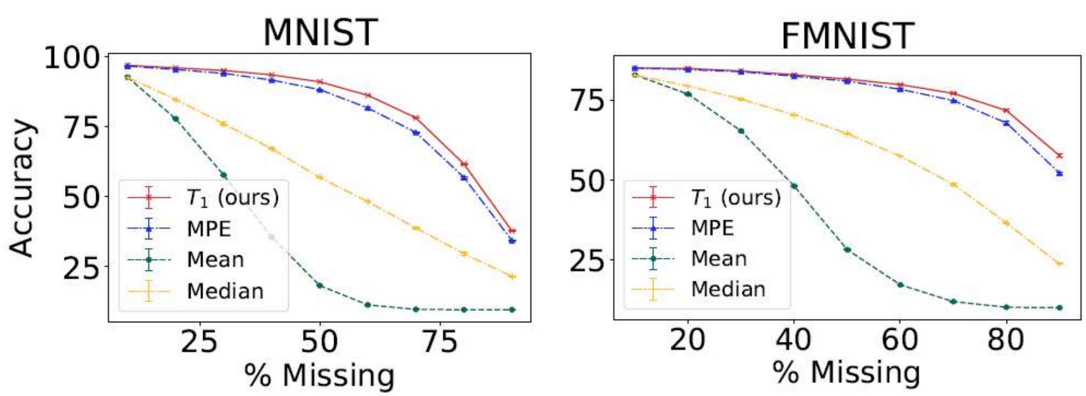
Can we solve these hard problems?

What to expect of classifiers?

- Missing features at prediction time
- What is expected prediction of F(x) in P(x)?

$$E_{\mathcal{F},P}(\mathbf{y}) = \mathop{\mathbb{E}}_{\mathbf{m} \sim P(\mathbf{M}|\mathbf{y})} \left[\mathcal{F}(\mathbf{y}\mathbf{m}) \right]$$

M: Missing features
y: Observed Features

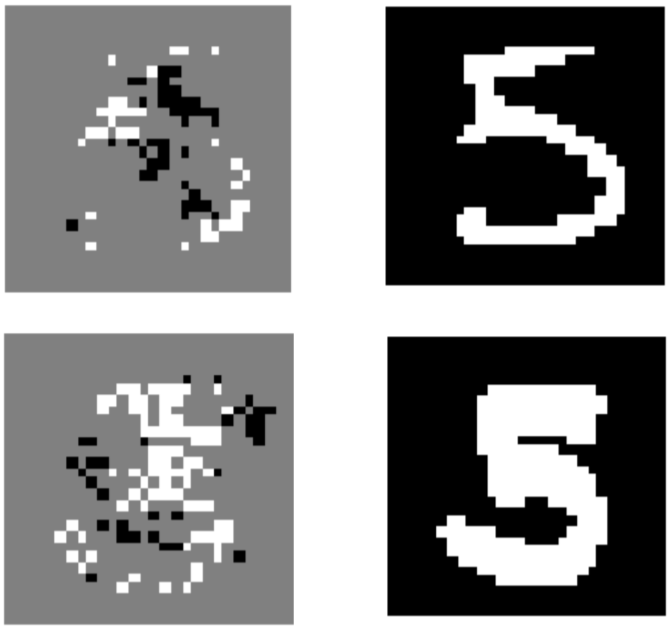


Explaining classifiers on the world

If the world looks like P(x),

then what part of the data is *sufficient* for

F(x) to make the prediction it makes?



Conclusions

Pure Logic **Probabilistic World Models** **Pure Learning**

Bring high-level representations, general knowledge, and efficient high-level reasoning to the world of probability

Bring back models of the world, supporting new tasks, and reasoning about what we have learned, without compromising learning performance

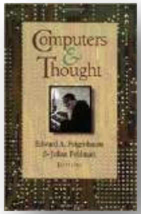
Conclusions

- There is a lot of value in working on pure logic, pure learning
- But we can do more
  by finding a synthesis, a confluence

- In another 56 years:

  Let’s get rid of this false dilemma…

References



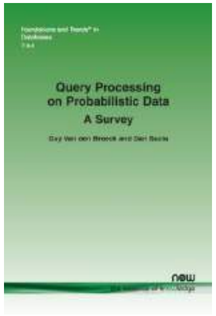
- **Computers and thought**
  Feigenbaum, E. A., & Feldman, J. (1963). McGraw-Hill.
- **Cards example of high-level reasoning**
  Guy Van den Broeck. Towards High-Level Probabilistic Reasoning with Lifted Inference, *In Proceedings of the AAAI Spring Symposium on KRR*, 2015.
- **Exchangeability as a source of tractability**
  Mathias Niepert and Guy Van den Broeck. Tractability through exchangeability: A new perspective on efficient probabilistic inference, *In Proceedings of the 28th AAAI Conference on Artificial Intelligence, AAAI Conference on Artificial Intelligence*, 2014.

References

- **FO^2 liftability theorem**
  Guy Van den Broeck, Wannes Meert and Adnan Darwiche. Skolemization for weighted first-order model counting, *In Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 2014.
- **Intractability of FO^3**
  Paul Beame, Guy Van den Broeck, Eric Gribkoff and Dan Suciu. Symmetric Weighted First-Order Model Counting, *In Proceedings of the 34th ACM Symposium on Principles of Database Systems (PODS)*, 2015.
- **Lifted learning**
  Jan Van Haaren, Guy Van den Broeck, Wannes Meert and Jesse Davis. Lifted Generative Learning of Markov Logic Networks, *In Machine Learning*, volume 103, 2015.

References

- **Confluences of ideas**
  Life in the Fast Lane: Viewed from the Confluence Lens. George Varghese, SIGCOMM CCR, 2015.
- **Probabilistic logic programming**
  Jonas Vlasselaer, Guy Van den Broeck, Angelika Kimmig, Wannes Meert and Luc De Raedt. Tp-Compilation for Inference in Probabilistic Logic Programs, *In International Journal of Approximate Reasoning*, 2016.
- **Probabilistic databases**
  Guy Van den Broeck and Dan Suciu. Query Processing on Probabilistic Data: A Survey, Foundations and Trends in Databases, Now Publishers, 2017.



- **Weighted model integration**
  Vaishak Belle, Andrea Passerini and Guy Van den Broeck. Probabilistic Inference in Hybrid Domains by Weighted Model Integration, *In Proceedings of 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

References

- **Probabilistic circuits**
  Antonio Vergari, Nicola Di Mauro and Guy Van den Broeck. Tractable Probabilistic Models, UAI Tutorial, 2019.
- **Logistic circuits**
  Yitao Liang and Guy Van den Broeck. Learning Logistic Circuits, *In Proceedings of the 33rd Conference on Artificial Intelligence (AAAI)*, 2019.
- **What to expect of classifiers?**
  Pasha Khosravi, Yitao Liang, YooJung Choi and Guy Van den Broeck. What to Expect of Classifiers? Reasoning about Logistic Regression with Missing Features, *In Proceedings of the ICML Workshop on Tractable Probabilistic Modeling (TPM)*, 2019.
  & unpublished work in progress

Thank You Messages

- IJCAI and the awards committee
- Luc De Raedt
- Stuart Russell, Dan Suciu, Kristian Kersting, David Poole, Dan Roth, George Varghese, Rina Dechter, Lise Getoor, Dan Olteanu
- Kurt Driessens, Wannes Meert, Arthur Choi
- The generousity of my many collaborators

  Mathias Niepert, Jesse Davis, Siegfried Nijssen, etc.
- Irma and my family and friends

Thank You Messages

- My amazing students & StarAI lab @ UCLA

- My “communities”: SRL, StarAI, Probabilistic (Logic) Programming, TPM, KC, …

- Funding

DARPA fwo NSF

NEC intel facebook research

Thank You All