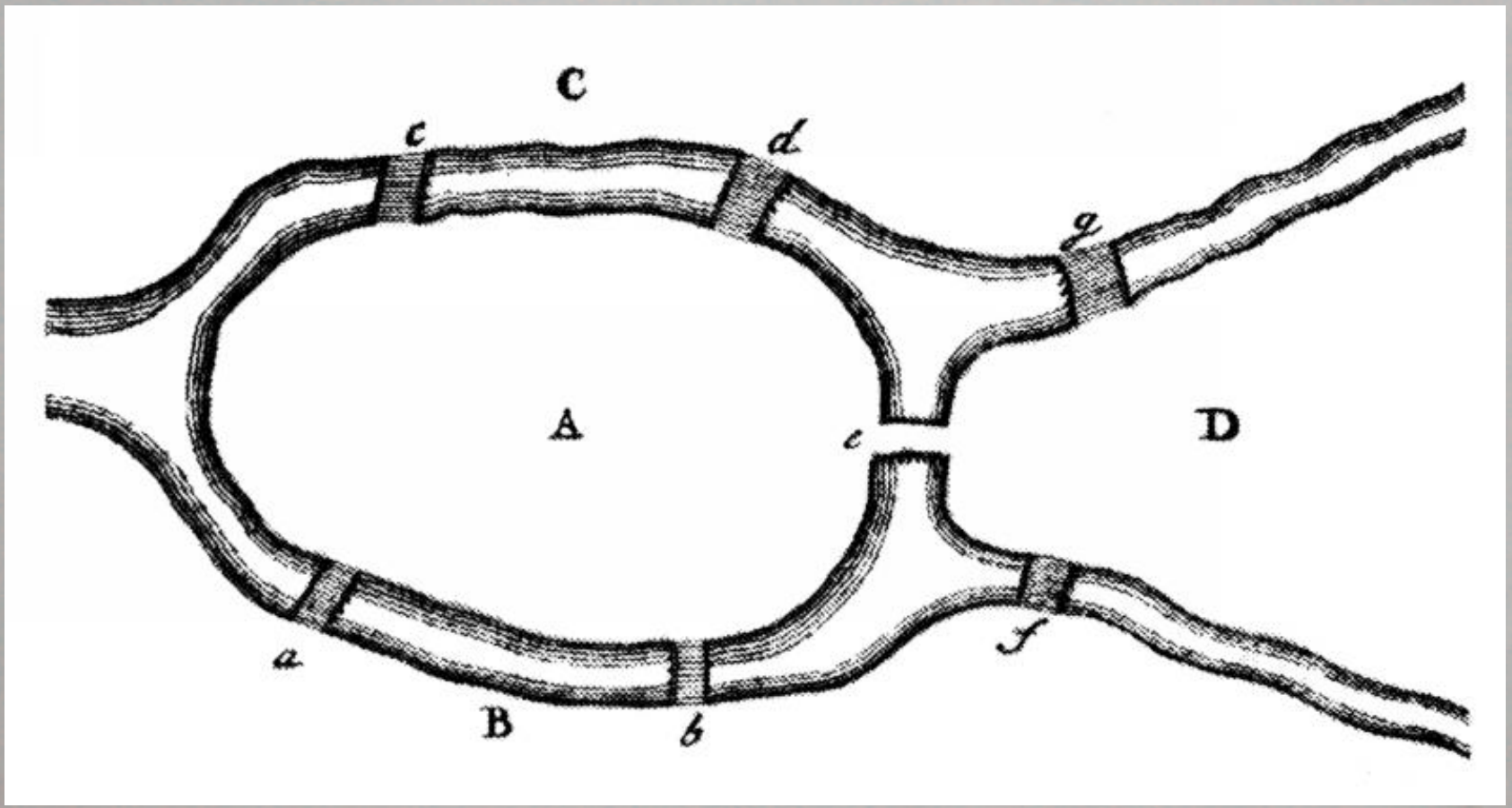
A Tale of Two Cultures: Phase Transitions in Physics and Computer Science

Cristopher Moore
University of New Mexico
and the Santa Fe Institute

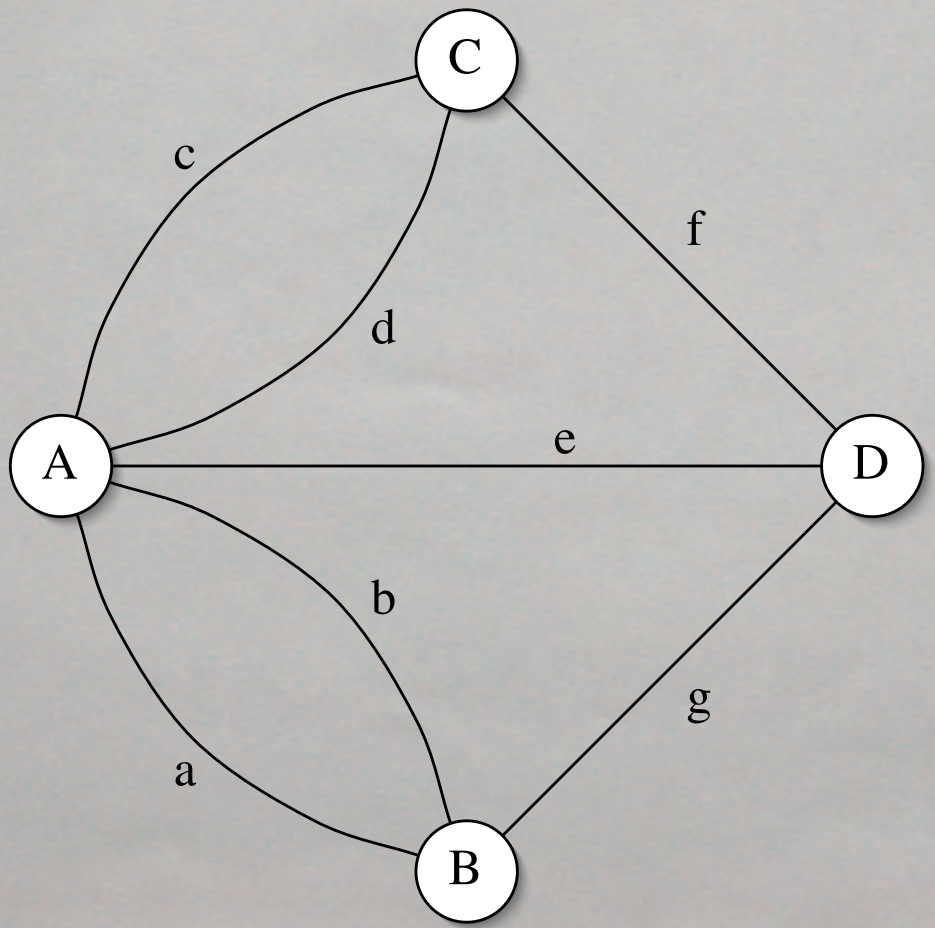
Computational Complexity

- Why are some problems qualitatively harder than others?



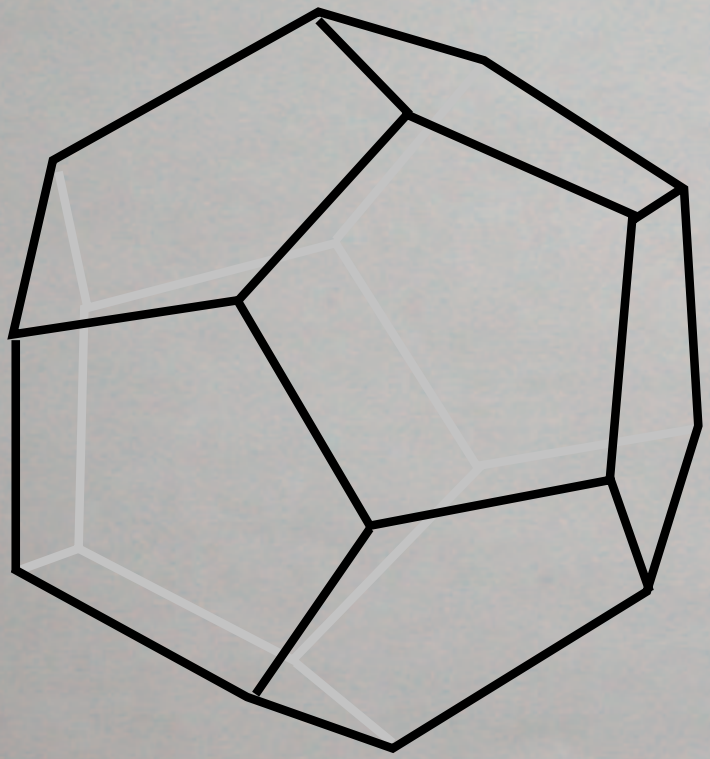
Computational Complexity

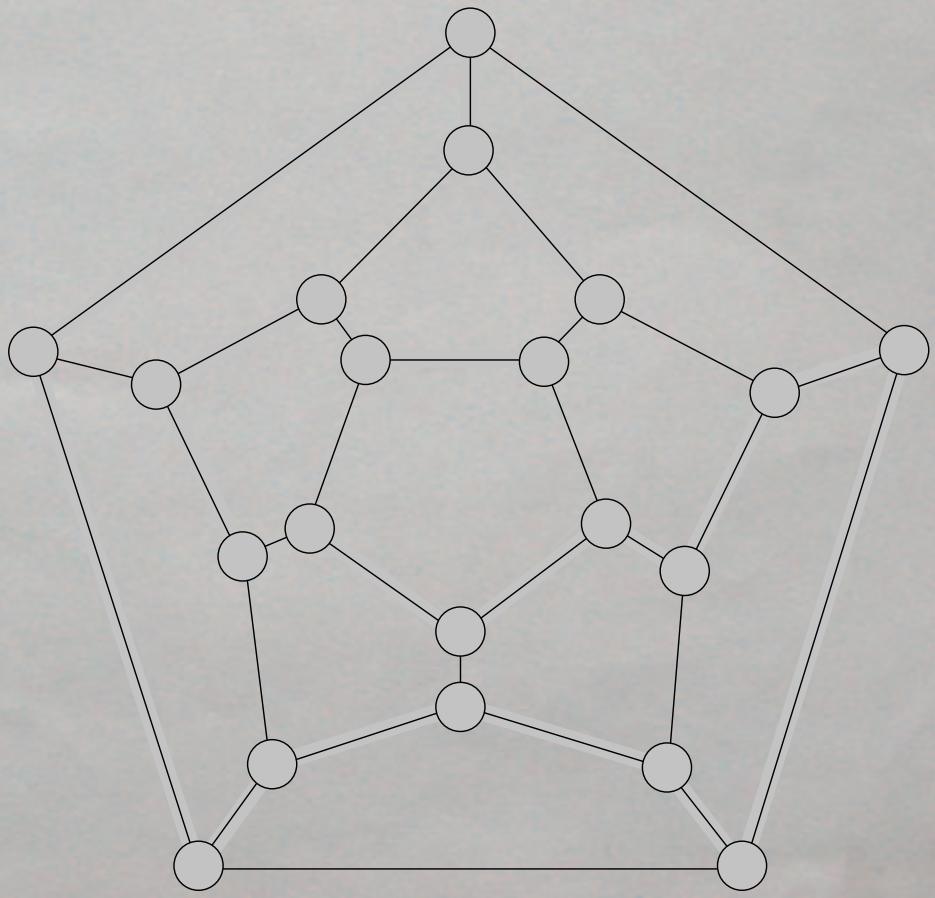
- A simple insight: at most 2 vertices can have odd degree, so no tour is possible!



Computational Complexity

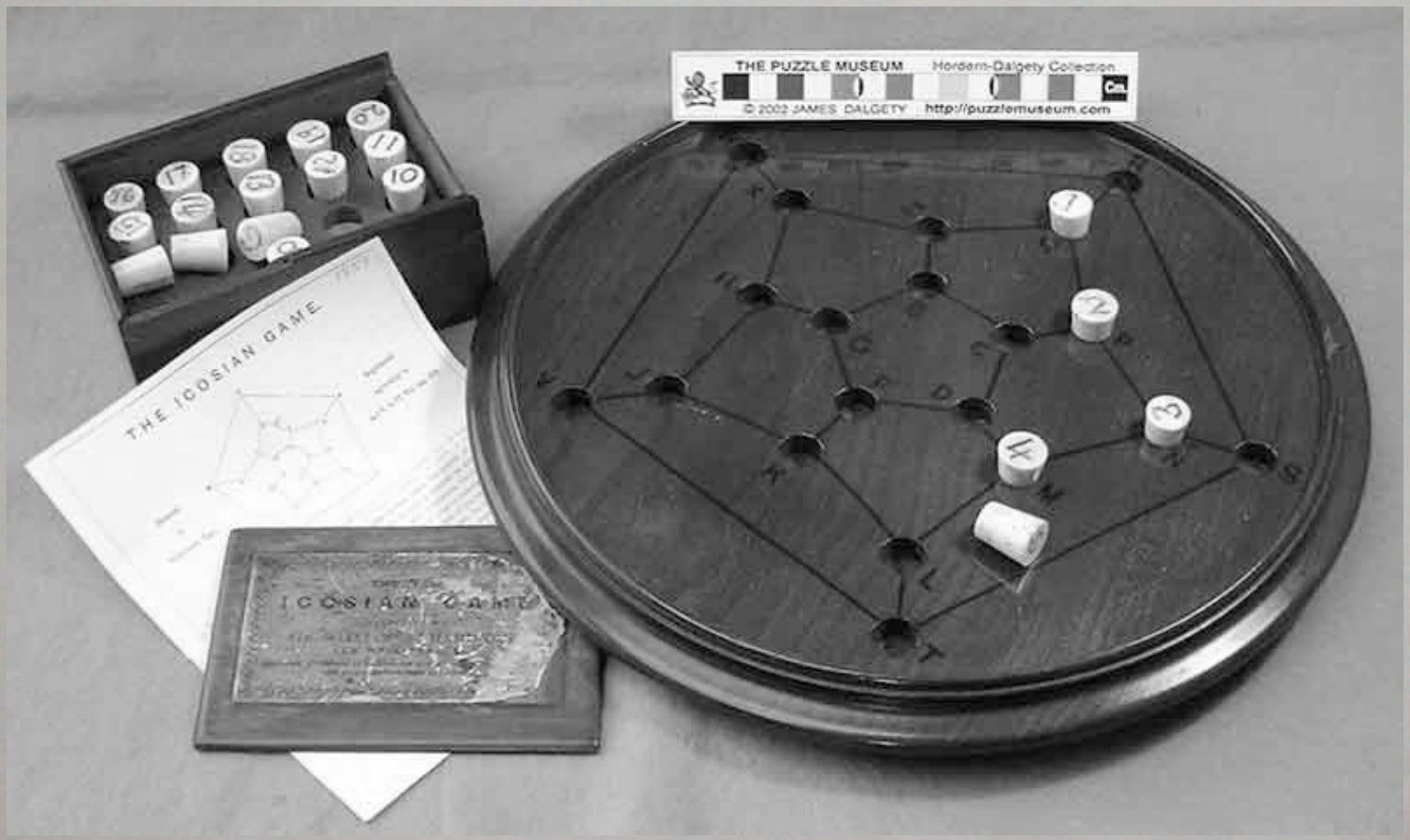
- What if we want to visit every vertex, instead of every edge?





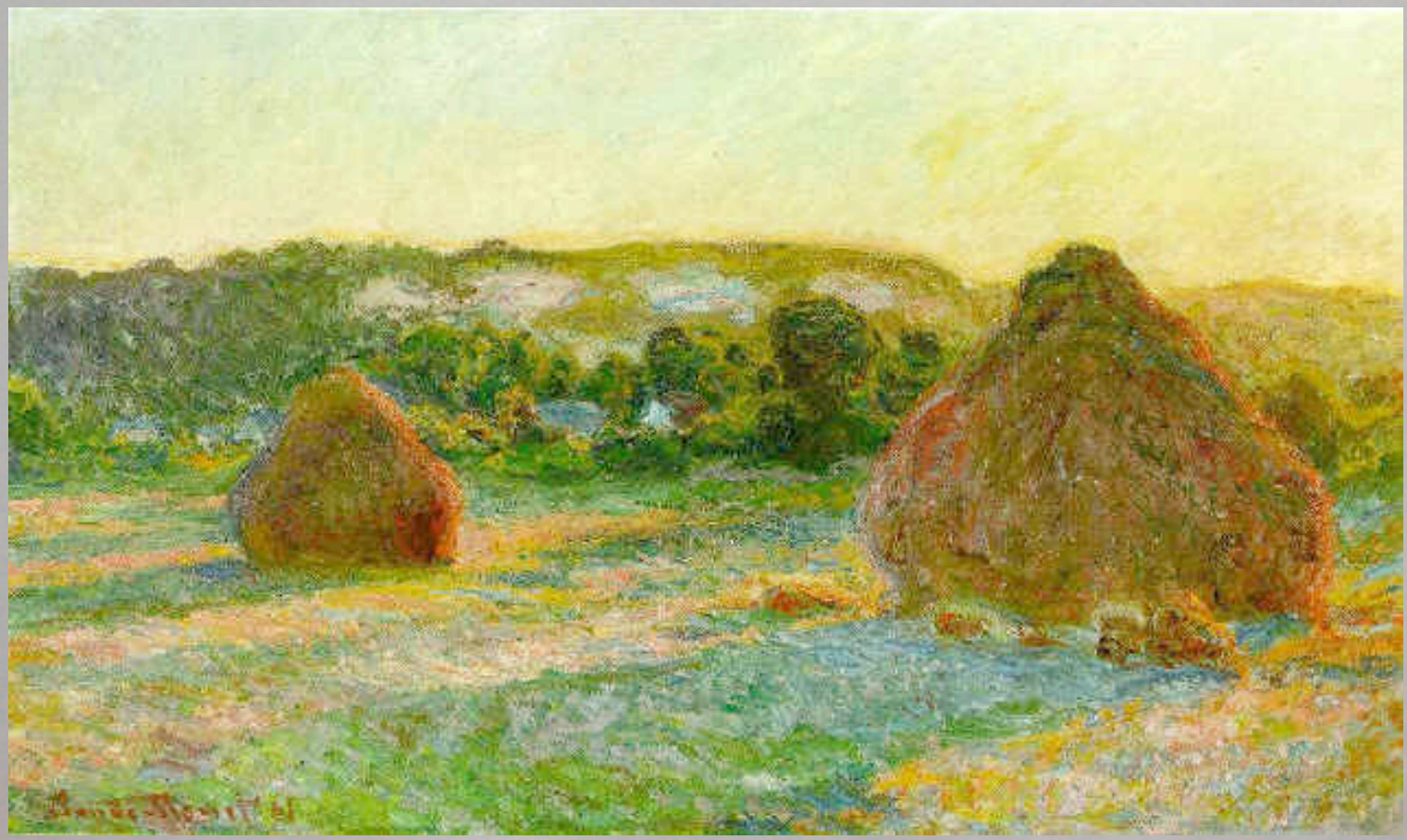
Computational Complexity

- As far as we know, the only way to solve this problem is (essentially) exhaustive search!

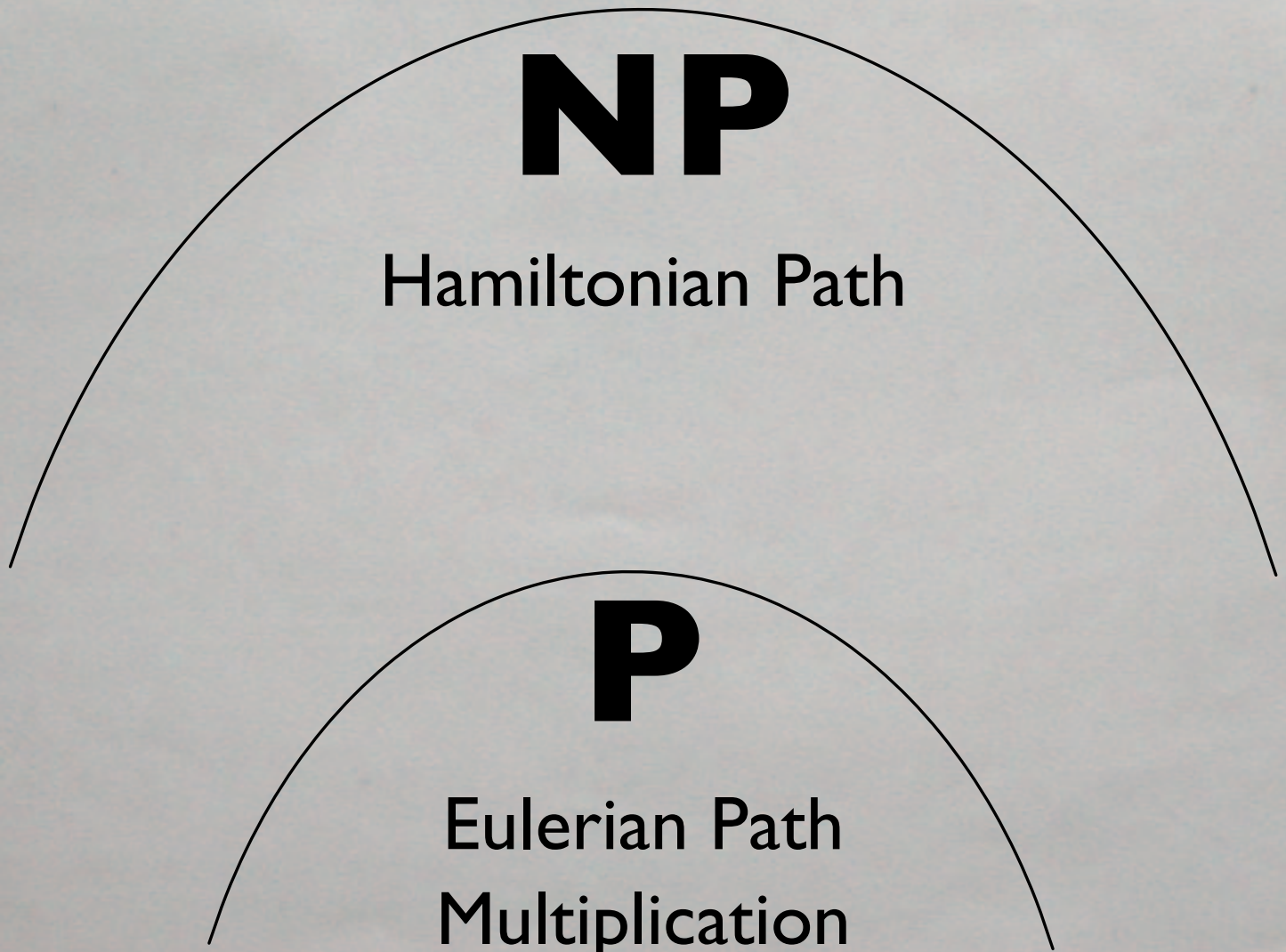


Needles in Haystacks

- **P**: we can find a solution efficiently
- **NP**: we can *check* a solution efficiently

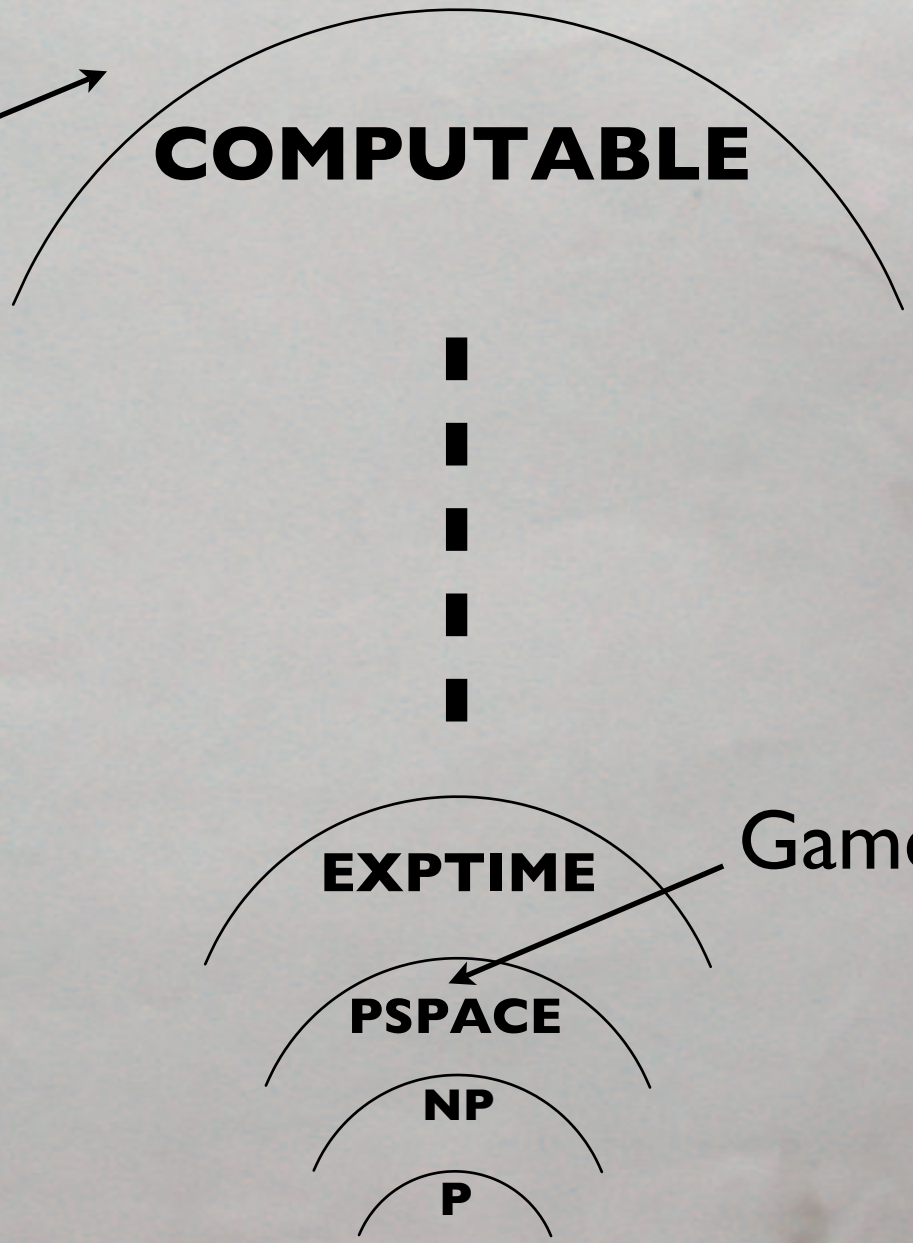


Complexity Classes



An Infinite Hierarchy

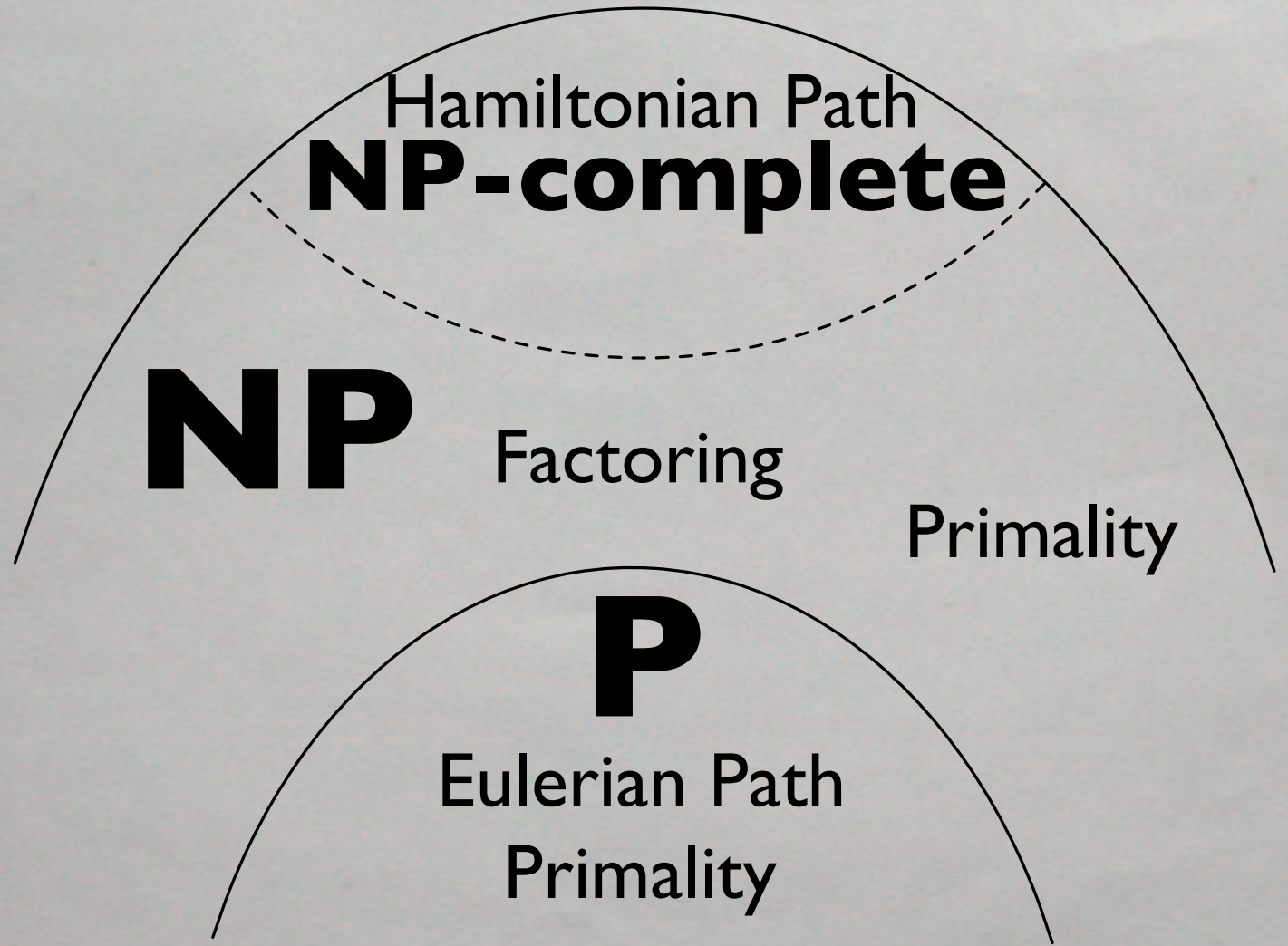
Turing's Halting Problem



"Computers play the same role in complexity that clocks, trains and elevators play in relativity."

– Scott Aaronson

The Hardest of Them All



Some problems in **NP** capture the entire class!

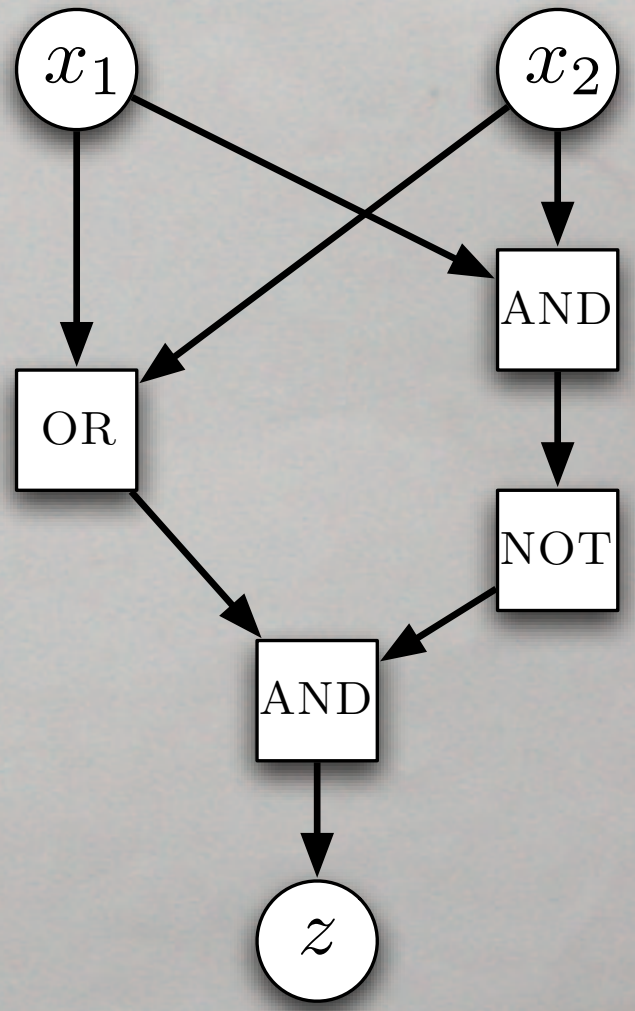
If we can solve any of them efficiently, then **P=NP.**

Satisfying a Circuit

Any program that tests solutions (e.g. paths) can be "compiled" into a Boolean circuit

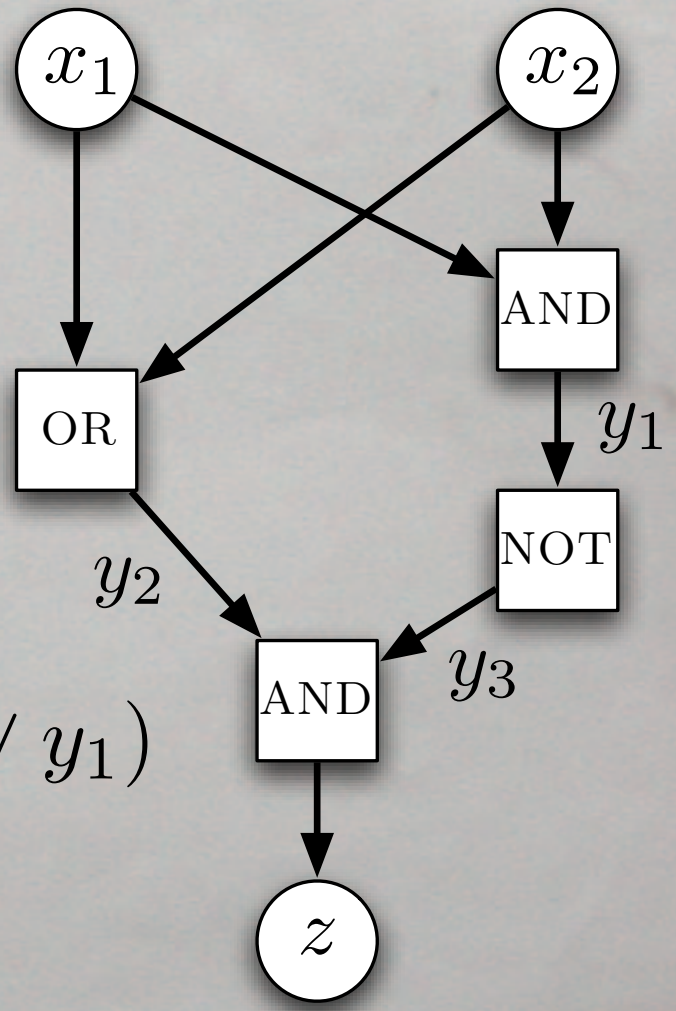
The circuit outputs "true" if an input solution works

Is there a set of values for the inputs that makes the output true?



From Circuits to Formulas

The condition that each AND or OR gate works, and the output is "true," can be written as a Boolean formula:



$$(x_1 \vee \overline{y}_1) \wedge (x_2 \vee \overline{y}_1) \wedge (\overline{x}_1 \vee \overline{x}_2 \vee y_1) \wedge \cdots \wedge z \ .$$

3-SAT

- Our first NP-complete problem!
- Given a set of *clauses* with 3 variables each,

$$(x_1 \vee \overline{x}_2 \vee x_3) \wedge (x_2 \vee x_{17} \vee \overline{x}_{293}) \wedge \cdots$$

  does a set of truth values for the x_i exist such that all the clauses are satisfied?

- k-SAT (k variables per clause) is NP-complete for $k \geq 3$.

If 3-SAT Were Easy...

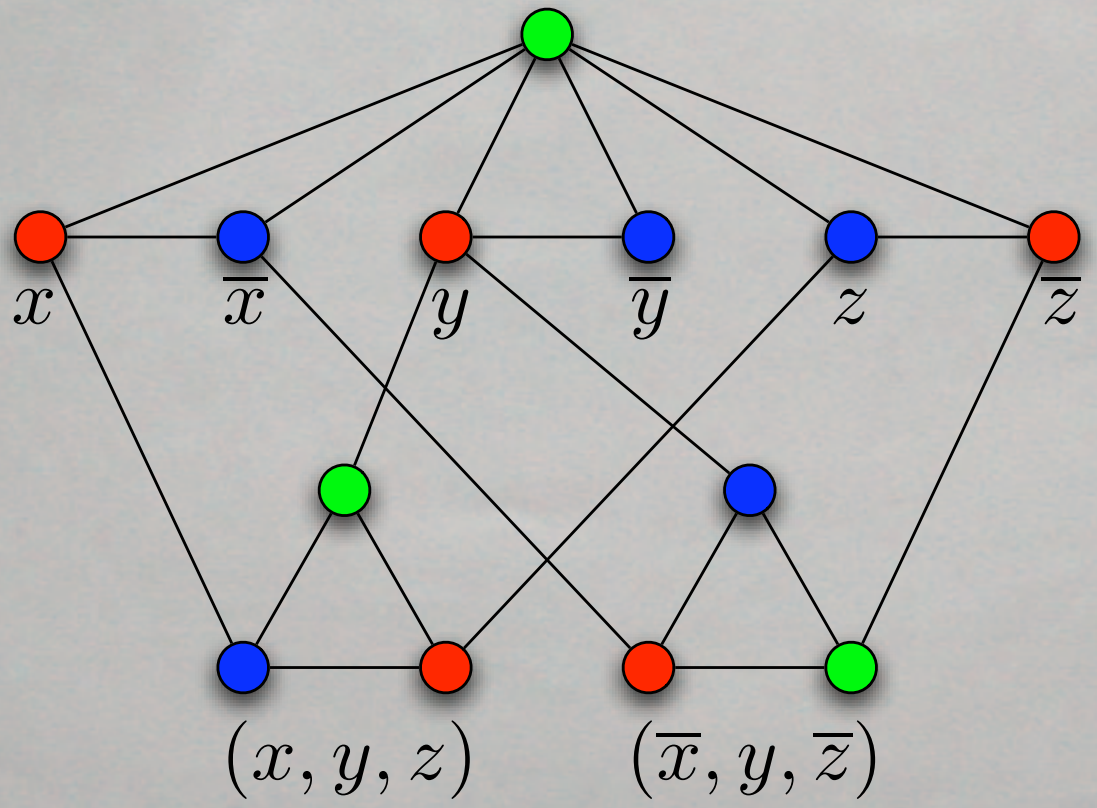
- ...we could convert any problem in NP to a circuit that tests solutions
- ...and convert that circuit to a 3-SAT formula which is satisfiable if a solution exists
- ...and use our efficient algorithm for 3-SAT to solve it!
- So, if 3-SAT is in **P**, then all of **NP** is too, and **P=NP!**

Graph Coloring

Given a set of countries and borders between them, what is the smallest number of colors we need?

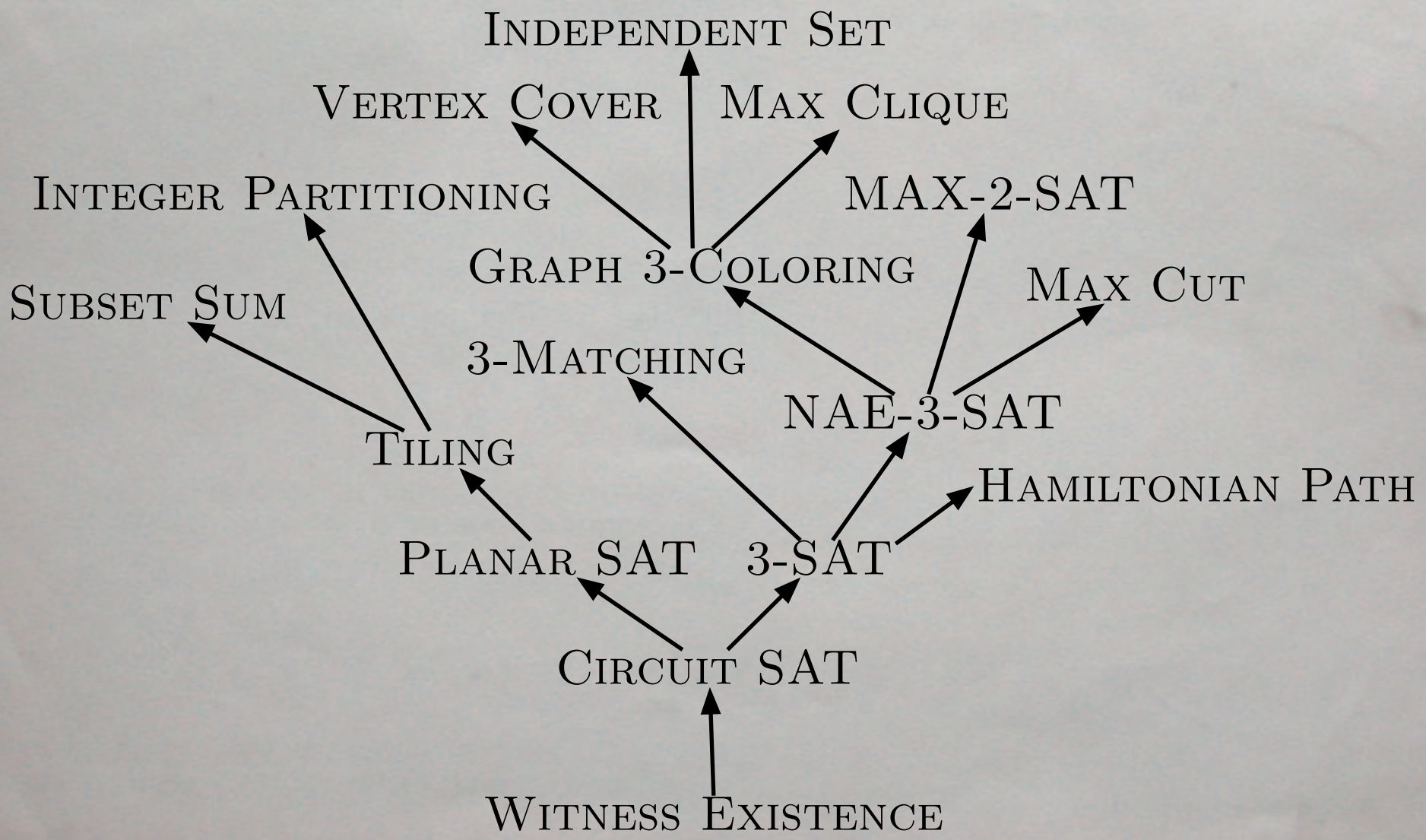
From SAT to Coloring

- "Gadgets" enforce constraints:



- Graph 3-Coloring is NP-complete
- Graph 2-Coloring is in **P** (why?)

And so on...


Independent Set
Vertex Cover
Max Clique
Integer Partitioning
MAX-2-SAT
Graph 3-Coloring
Subset Sum
Max Cut
3-Matching
NAE-3-SAT
Tiling
Hamiltonian Path
Planar SAT
3-SAT
Circuit SAT
Witness Existence

The Adversary

...designs problems that are as diabolically hard as possible, forcing us to solve them in the worst case. (Hated and feared by computer scientists.)

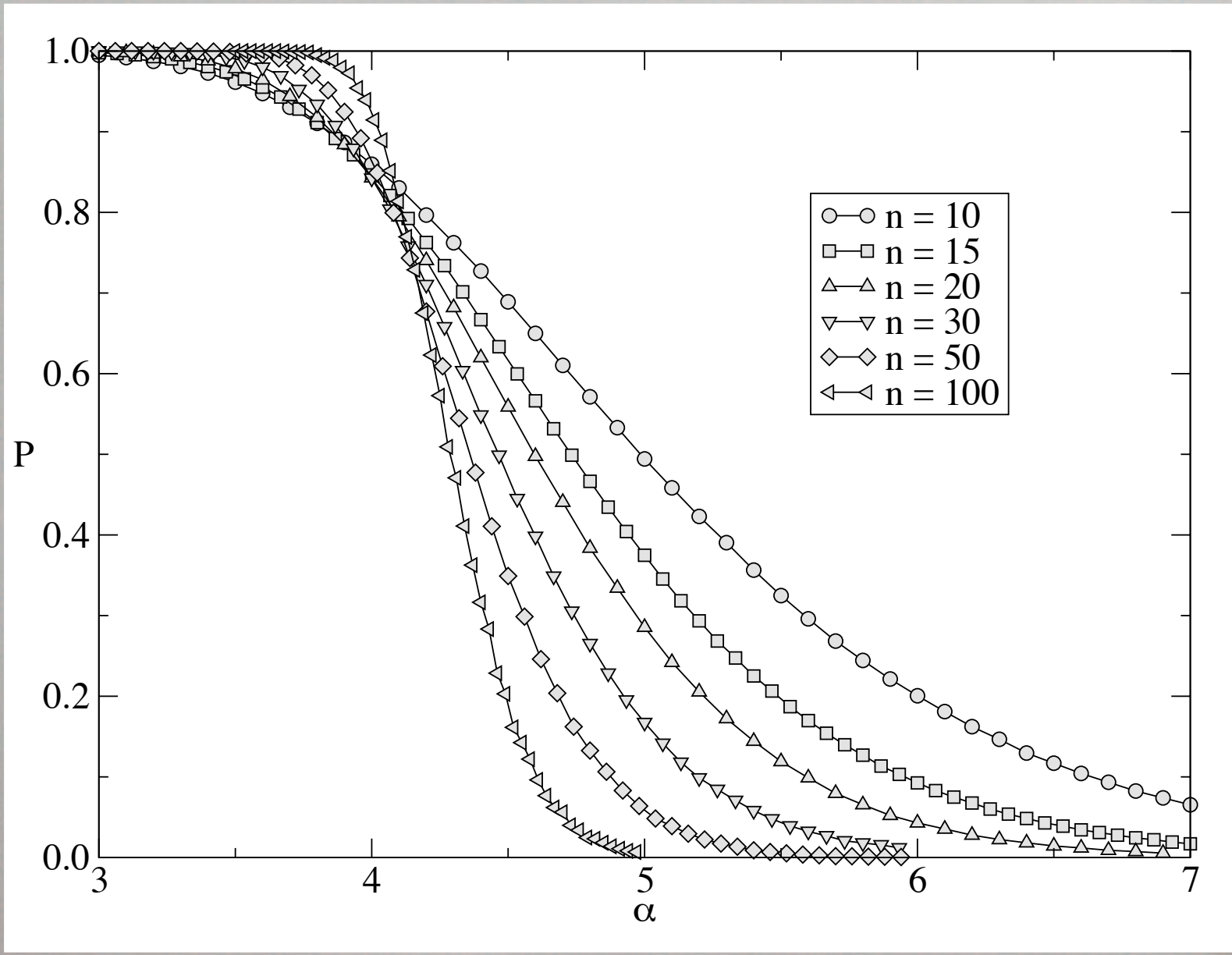
La Dame Nature

...asks questions whose answers are simpler and more beautiful than we have any right to imagine. (Loved and worshipped by physicists.)

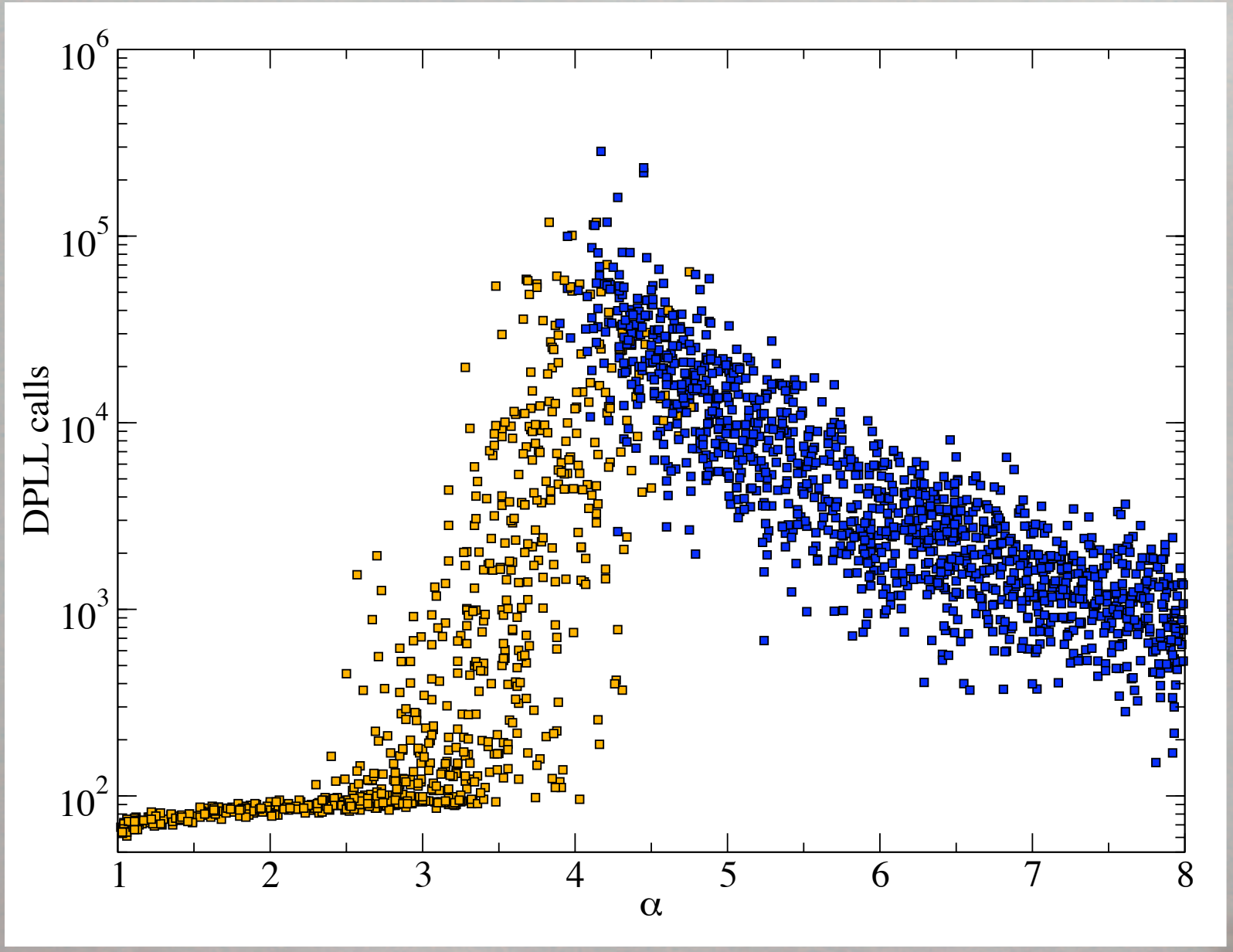
Random Problems

- A 3-SAT formula with n clauses, m variables
- Choose each clause randomly: $\binom{n}{3}$ possible triplets, negate each one with probability 1/2
- Precedents:
  - Random Graphs (Erdös-Rényi)
  - Statistical Physics: ensembles of disordered systems, e.g. spin glasses
- Sparse Case: $m = \alpha n$ for some density α

A Phase Transition



Search Times



An Upper Bound

- The *average* number of solutions $\mathrm{E}[X]$ is

$$2^n \left(\frac{7}{8}\right)^m = \left(2\left(\frac{7}{8}\right)^\alpha\right)^n$$

- This is exponentially small whenever

$$\alpha > \log_{8/7} 2 \approx 5.19$$

- But the transition is much lower, at $\alpha \approx 4.27$. What's going on?

A Heavy Tail

- In the range $4.27 < \alpha < 5.19$, the average number of solutions is exponentially large.
- Occasionally, there are exponentially many...
- ...but most of the time there are none!
- A classic "heavy-tailed" distribution
- Large average doesn't prove satisfiability!

Lower Bound #1

- Idea: track the progress of a simple algorithm!
- When we set variables, clauses disappear or get shorter:

$$\overline{x} \wedge (x \vee y \vee z) \Rightarrow (y \vee z)$$

- *Unit Clauses* propagate:

$$x \wedge (\overline{x} \vee y) \Rightarrow y$$

One Path Through the Tree

- If there is a unit clause, satisfy it. Otherwise, choose a random variable and give it a random value!
- The remaining formula is random for all *t*:

$$\frac{ds_3}{dt} = -\frac{3s_3}{1-t} \quad , \quad \frac{ds_2}{dt} = \frac{(3/2)s_3 - 2s_2}{1-t}$$

$$s_3(0) = \alpha \quad , \quad s_2(0) = 0$$

A Branching Process

- Each unit clause has on average λ children, where

$$\lambda = \frac{s_2}{1-t}$$

- When $\lambda > 1$, they proliferate and contradictions appear
- But if $\alpha < 8/3$, then $\lambda < 1$ throughout and the unit clauses stay manageable.

Constructive Methods Fail

- Fancier algorithms, harder math: $\alpha < 3.52$.
- But, for larger k, algorithmic methods are nowhere near the upper bound for k-SAT:

$$O\left(\frac{2^k}{k}\right) < \alpha < O(2^k)$$

- To close this gap, we need to resort to non-constructive methods.

Lower Bound #2

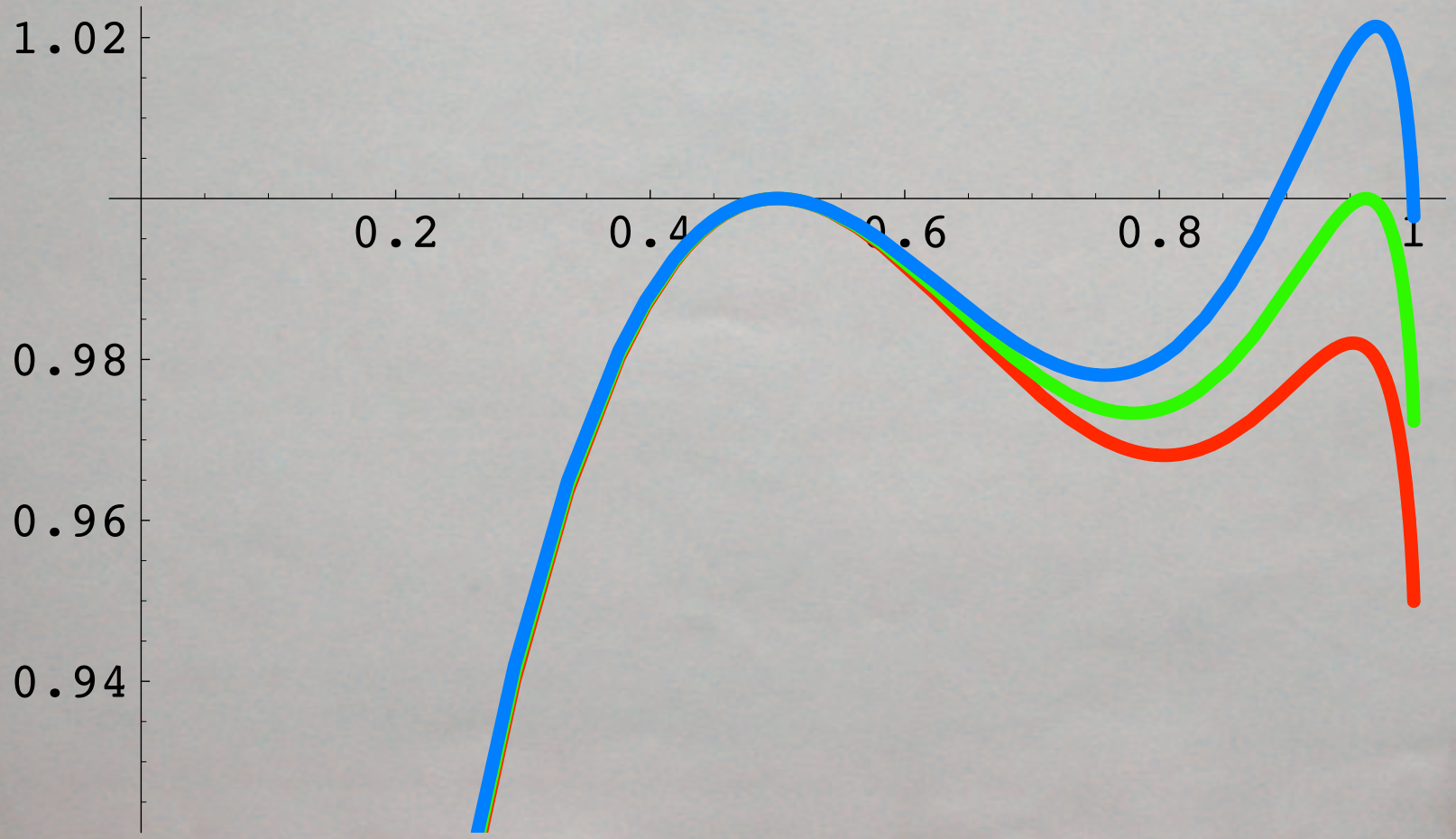
- Idea: bound the *variance* of the number of solutions.
- If X is a nonnegative random variable,

$$\Pr[X > 0] \geq \frac{\mathrm{E}[X]^2}{\mathrm{E}[X^2]}$$

- $\mathrm{E}[X]$ is easy; $\mathrm{E}[X^2]$ requires us to understand *correlations* between solutions.

A Function of Distance

- When the expected number of *pairs* of solutions is peaked at 1/2, most pairs are "independent" and the variance is small.



Determining the Threshold

- A series of results has narrowed the range for the transition in k-SAT to

$$2^k \ln 2 - O(k) < \alpha < 2^k \ln 2 - O(1)$$

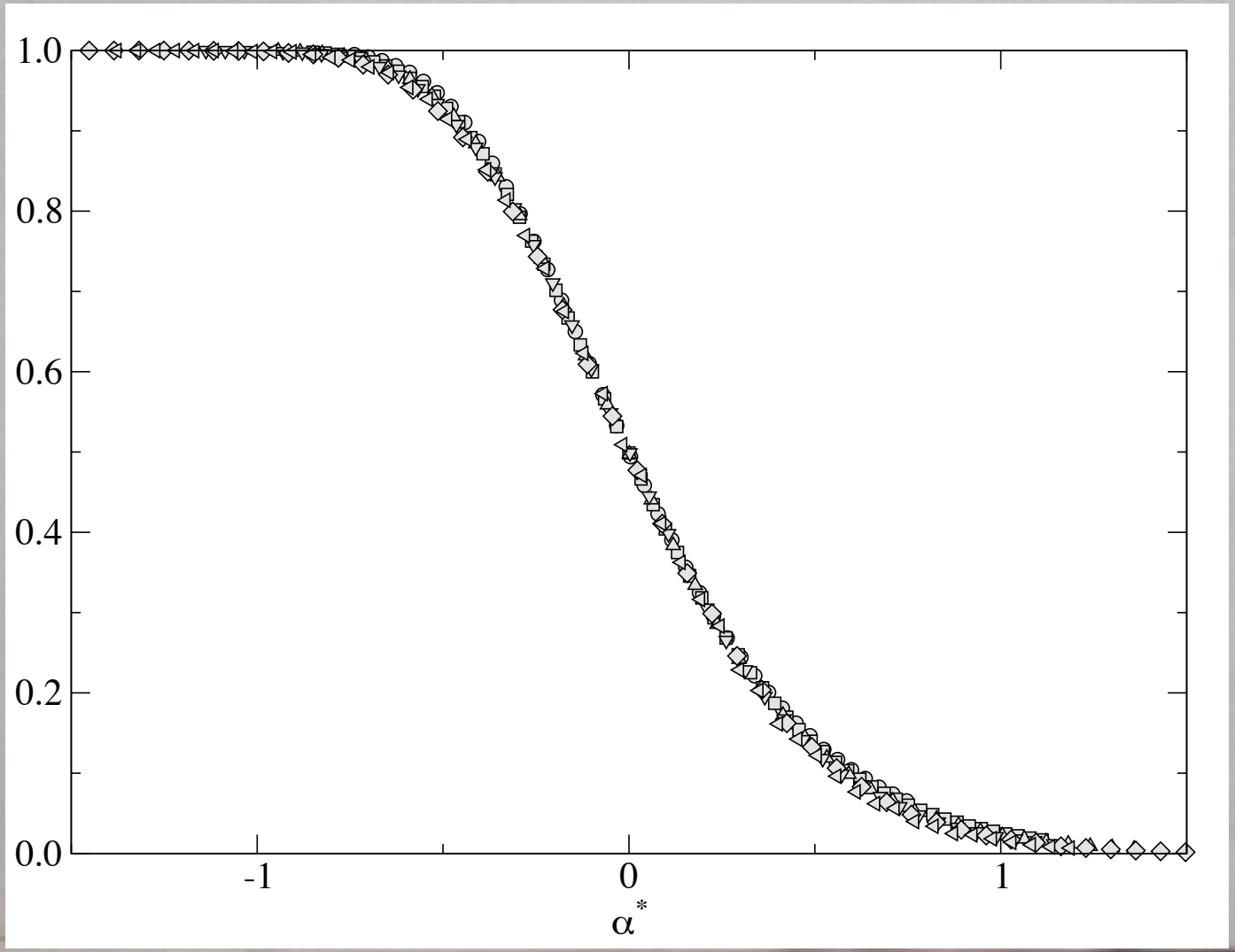
- Prediction from statistical physics:

$$2^k \ln 2 - O(1)$$

- Seems difficult to prove with current methods.

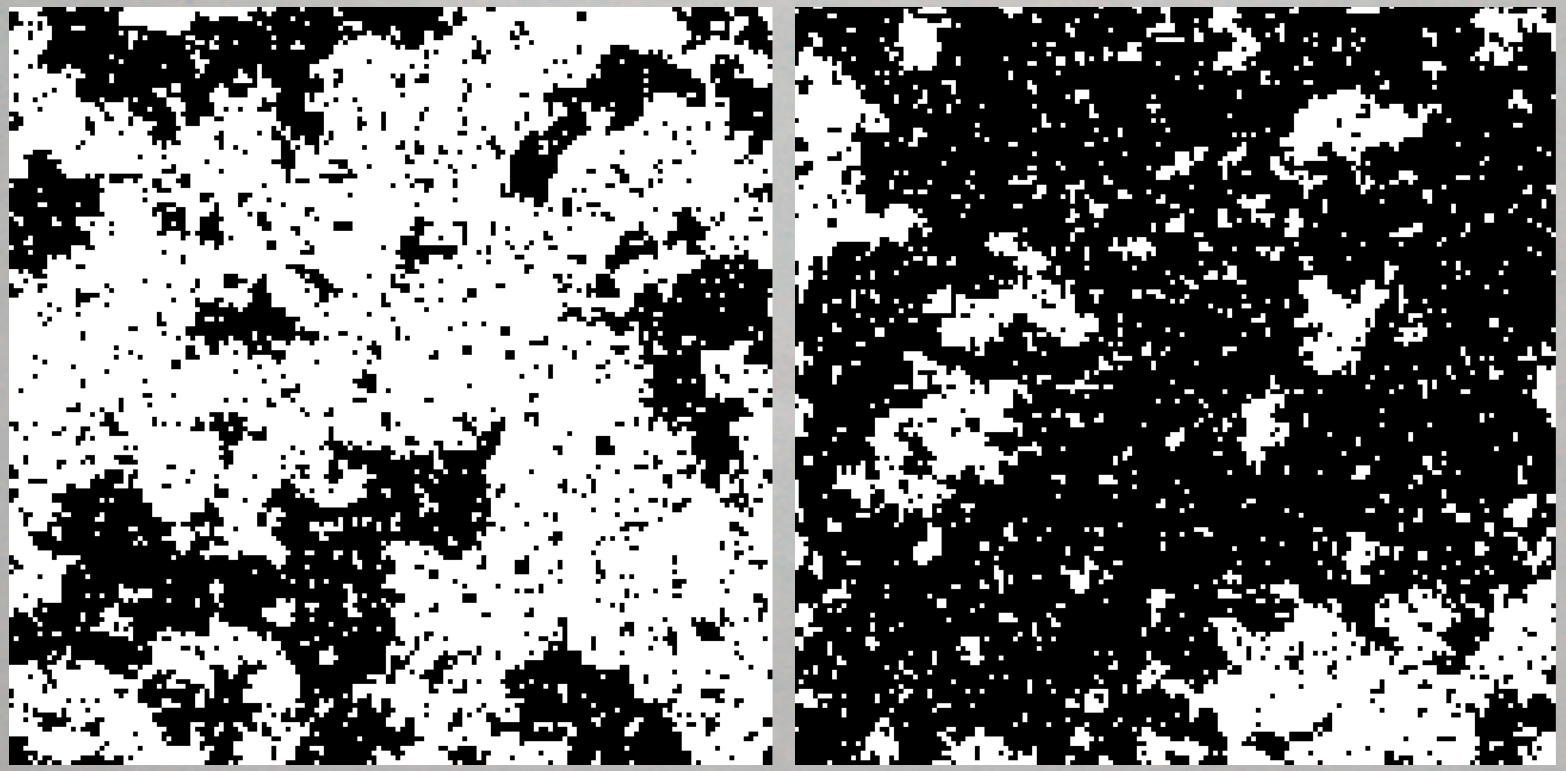
Scaling and Universality

- Rescaling α around the critical point causes different n to coincide. A universal function?



Clustering

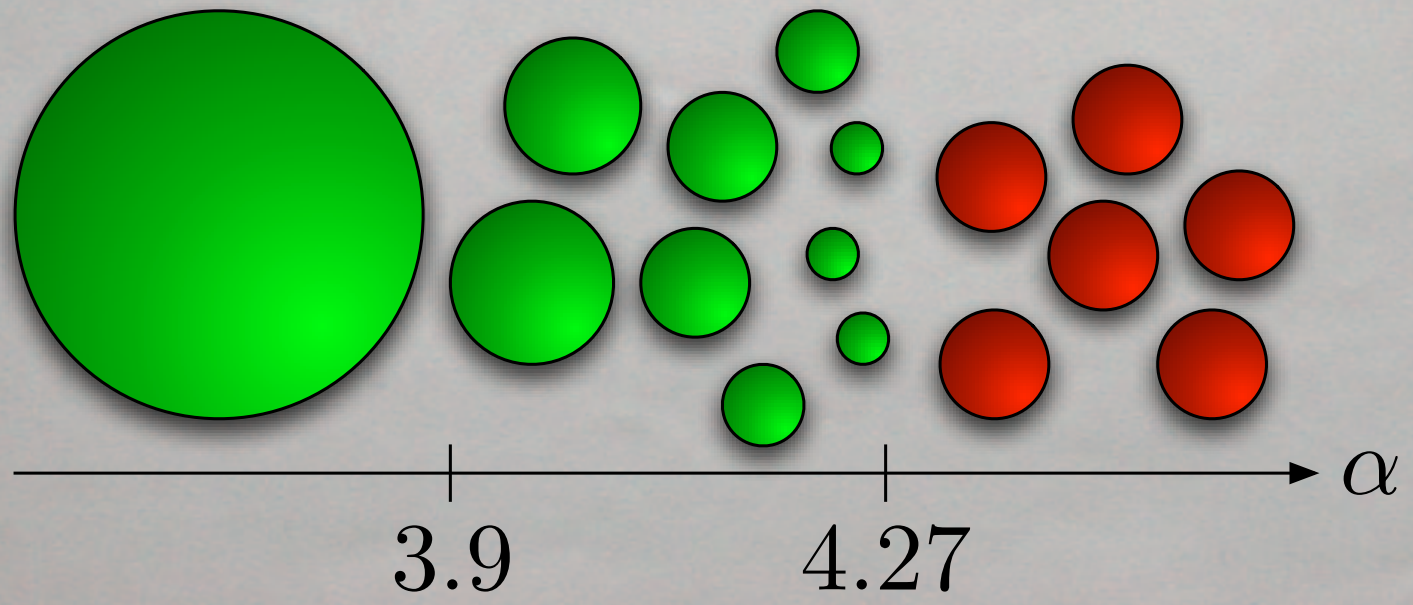
- Below the critical temperature, magnets have two *macrostates* (Gibbs measures)



- Glasses, and 3-SAT, have exponentially many!

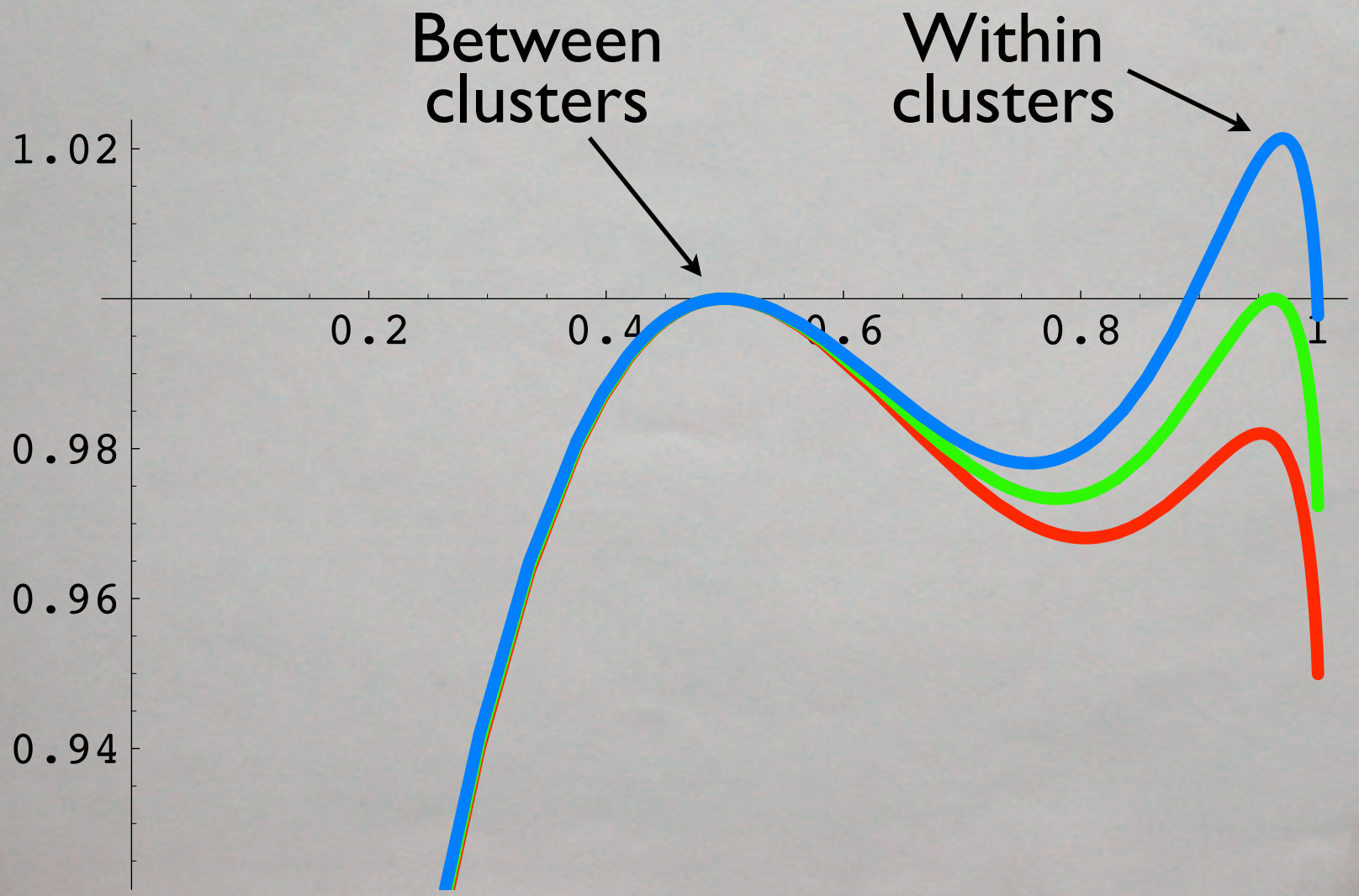
Clustering

- An idea from statistical physics: there is another transition, from a unified "cloud" of solutions to separate clusters.
- Is this why algorithms fail at $\alpha \sim 2^k / k$?



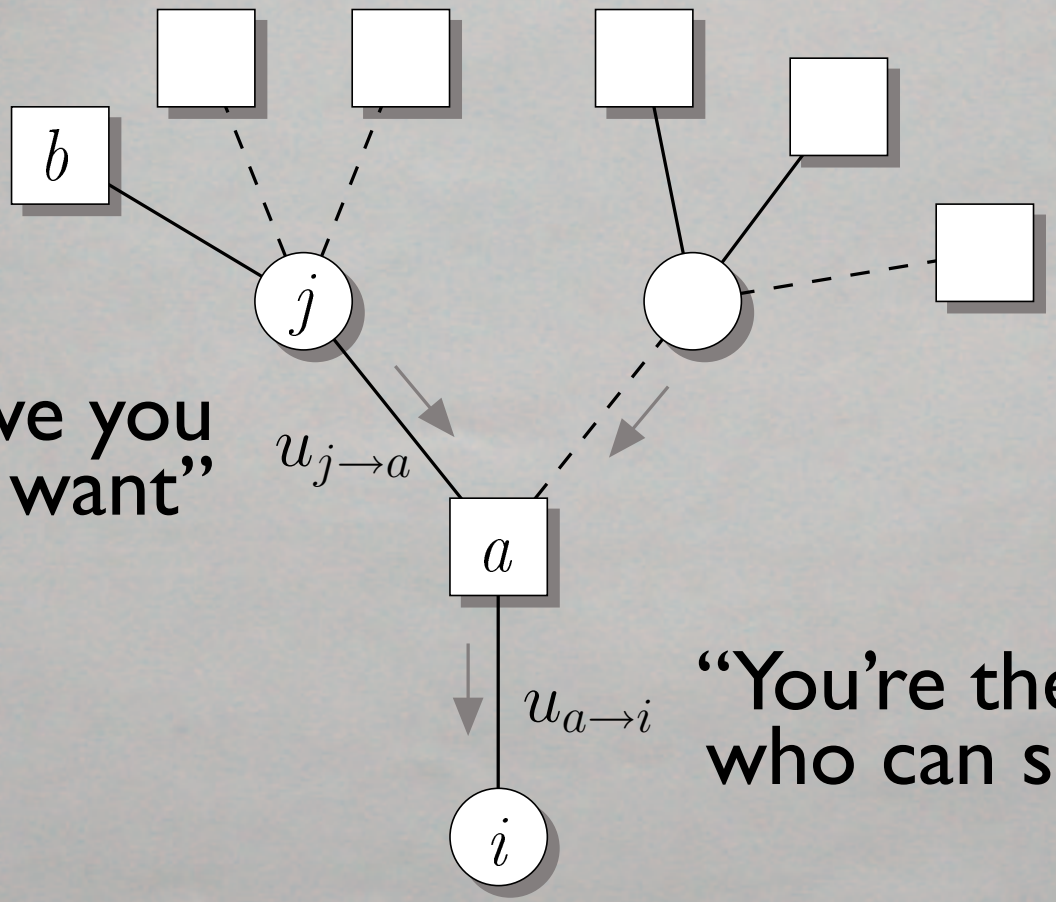
Distance Distributions

- Some rigorous evidence for clustering:



The Physicists' Algorithm

- A "message-passing" algorithm:

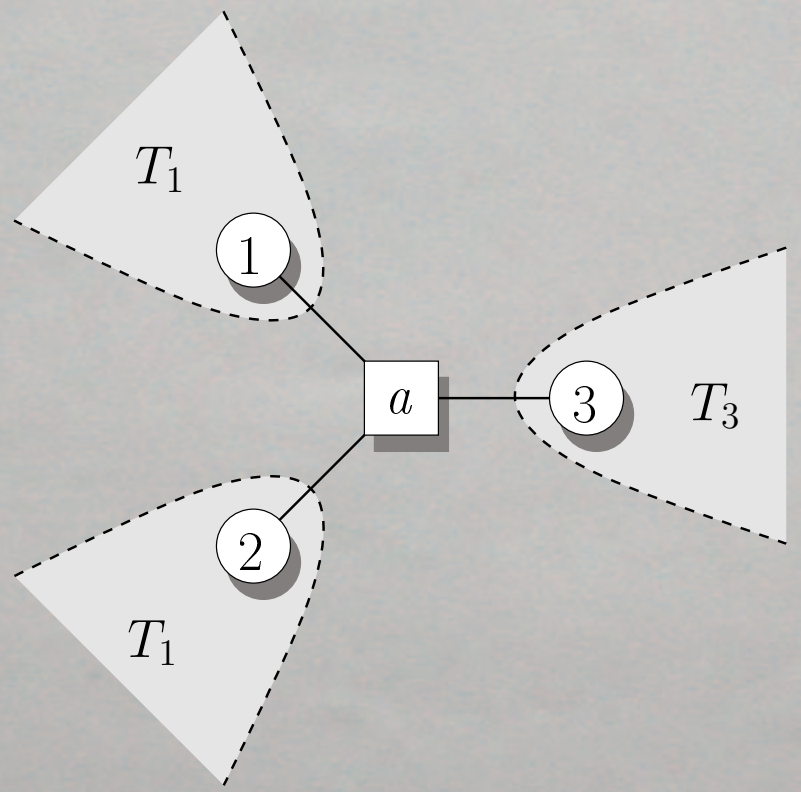


"I can't give you what you want"

"You're the only one who can satisfy me"

Why Does It Work?

- Random formulas are locally treelike.
- Assume the neighbors are independent:



- *Proving* this will take some very deep work.

Clay Mathematics Institute

Dedicated to increasing and disseminating mathematical knowledge

How to make $1,000,000 (and maybe $7,000,000)

Millennium Problems

- **P=NP**?
- Poincaré Conjecture
- Riemann Hypothesis
- Yang-Mills Theory
- Navier-Stokes Equations
- Birch and Swinnerton-Dyer Conjecture
- Hodge Conjecture

Millennium Problems

- **P=NP**?

Millennium Problems

- **P=NP**?
- Is it harder to find solutions than to check them?
- Question about the *nature of mathematical truth*
- ...and whether finding it requires as much creativity as we think.

What if P=NP?

- Better Traveling Salesmen, can pack luggage
- No Cryptography
- The entire *polynomial hierarchy* collapses, NEXP=EXP, etc.
- We can find (up to any reasonable length)
  - Proofs
  - Theories
  - *Anything* we can recognize.

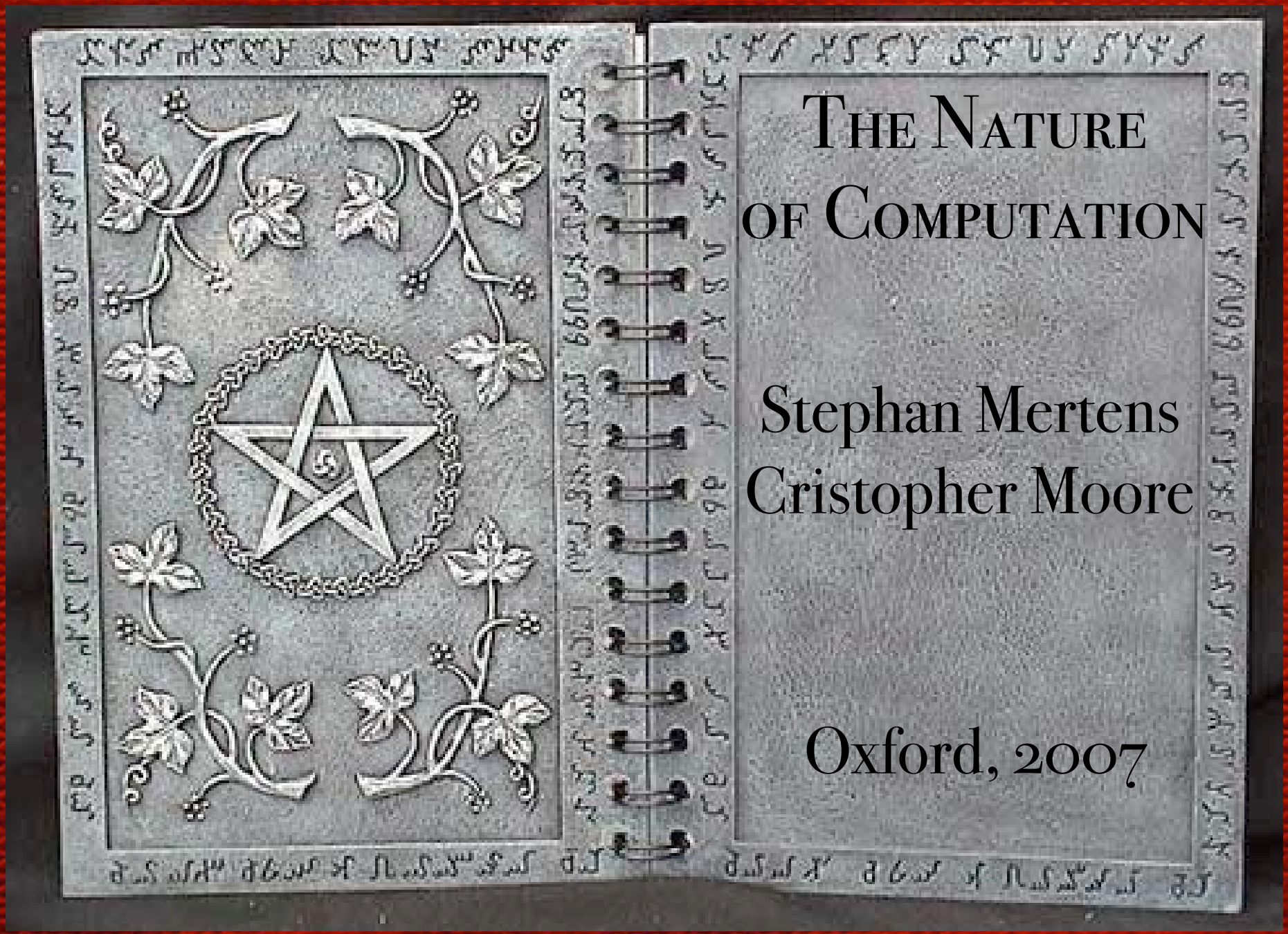
Gödel to Von Neumann

Let $\varphi(n)$ be the time it takes to decide whether a proof of length n exists. Gödel writes:

The question is, how fast does $\varphi(n)$ grow for an optimal machine. If there actually were a machine with, say, $\varphi(n) \sim n^2$, this would have consequences of the greatest magnitude. That is to say, it would clearly indicate that, despite the unsolvability of the *Entscheidungsproblem*, **the mental effort of the mathematician in the case of yes-or-no questions could be completely replaced by machines**. One would simply have to select an n large enough that, if the machine yields no result, there would then be no reason to think further about the problem.

P<NP if understanding matters.

Shameless Advertisement



Acknowledgments