Is evolution about learning?

Readings for today

- Valiant, L. G. (2009). Evolvability. Journal of the ACM (JACM), 56(1), 1-21.

Topics

- Entropic view of life
- Structure of a learnable problem

Entropic view of life

Let’s talk about life

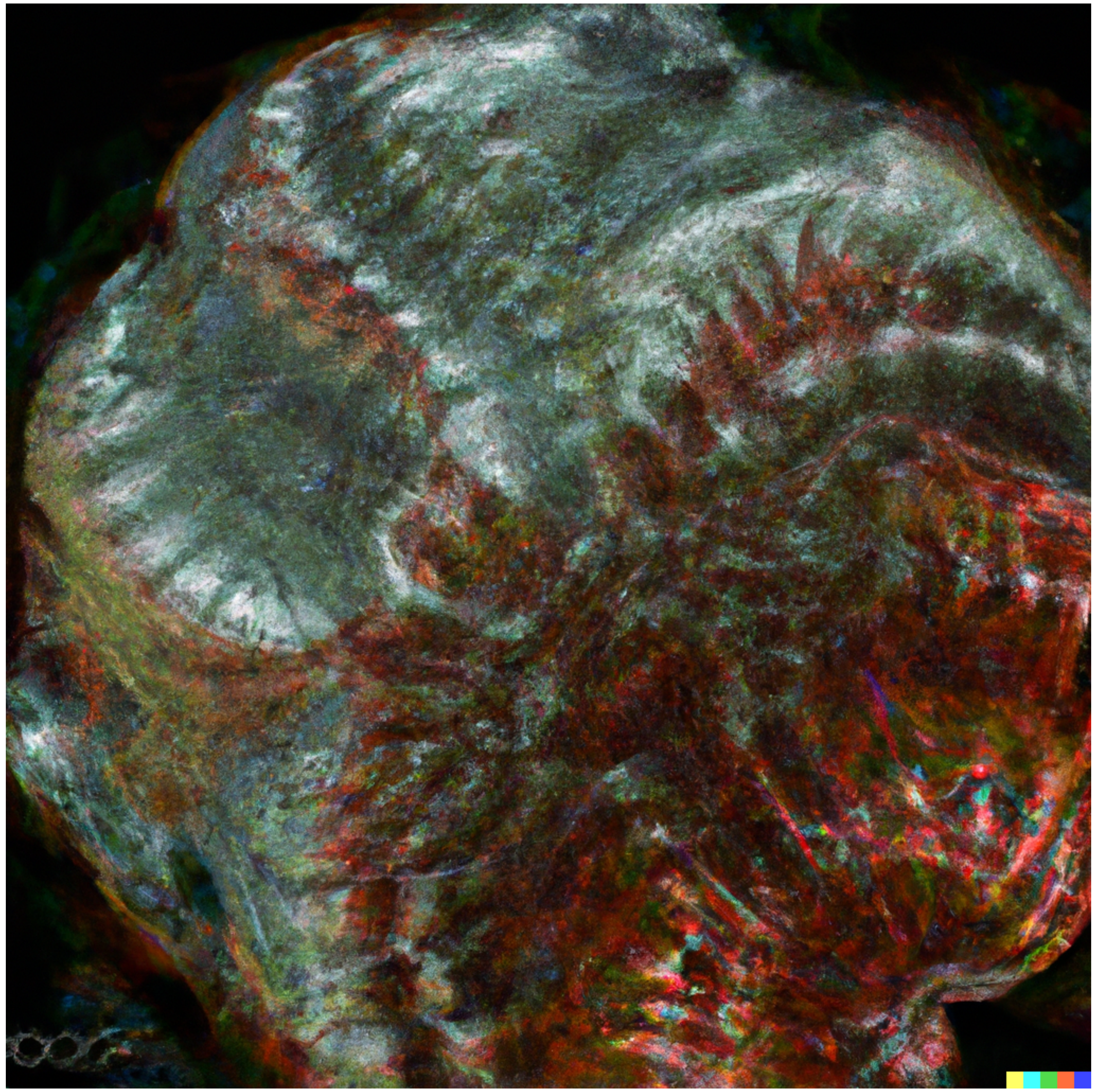
Question: What is life?

Answer: It’s complicated

- Cellular life
- Autocatalytic molecules
- Generation of proteins
- Reproduction
- Adaptation

Two basic properties:

- **Metabolism**: maintain their own structure by continually resynthesizing their own elements.
- **Replication**: Produce copies of themselves in the immediate vicinity.

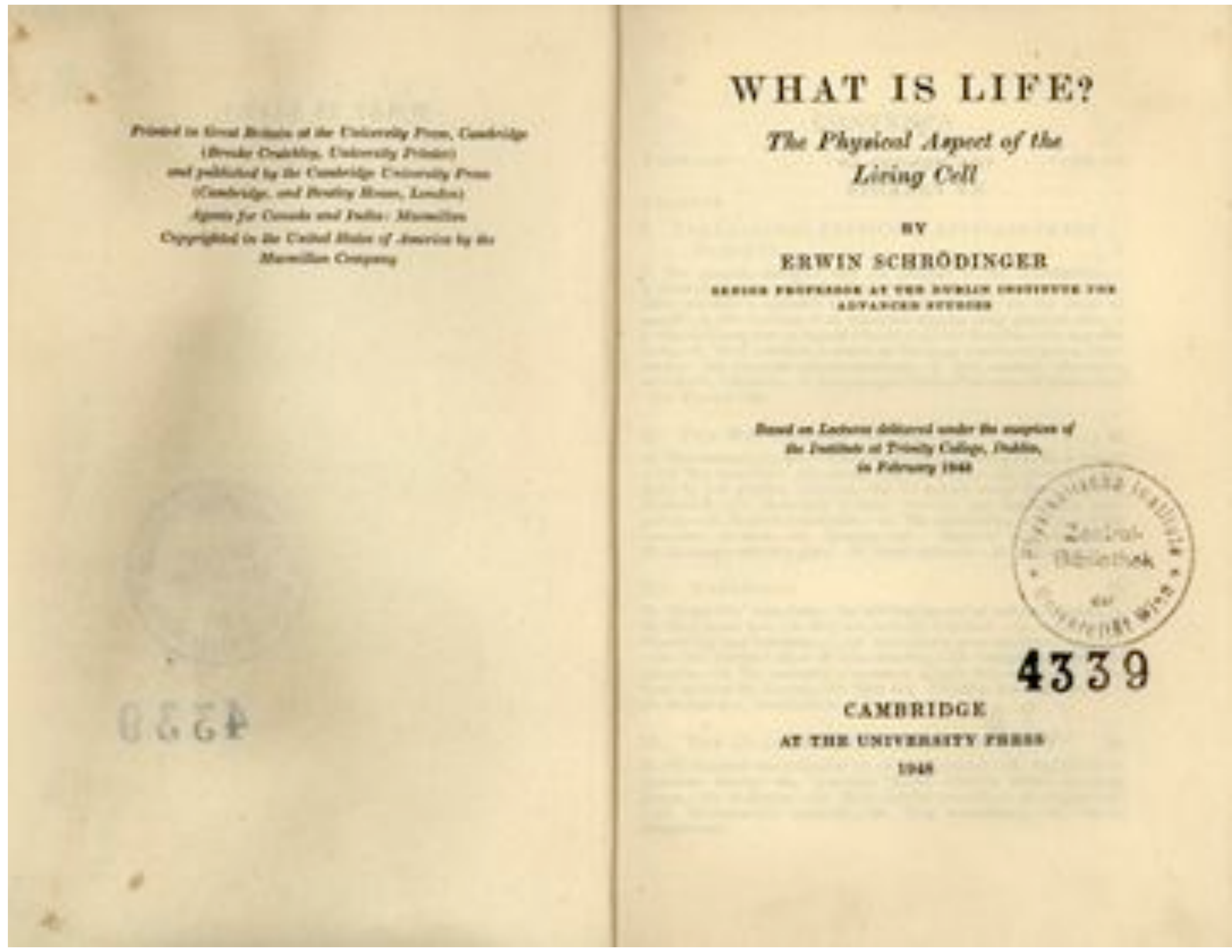


A better question

<u>Question</u>: What does life do?

<u>Answer</u>: “Order from order” - E. Schrödinger

- Life controls/reduces entropy of its environment.
- Manipulates local entropy so as to maintain molecular stability.



Entropy: Number of possible states that matter can take on.

$$S = -k_B \Sigma_i p_i \ln(p_i)$$

Entropy

Probability of state *i*

Boltzmann constant

Stochastic model of evolution

Goal: Maximize fitness of organism

- Random mutations (introduce change)
- Natural selection (select best fit)

Problem: If the process is a random search, then there are not enough seconds since the start of the universe to produce life as complex as what we have currently on earth.

NP-hard problem

P vs. NP complexity

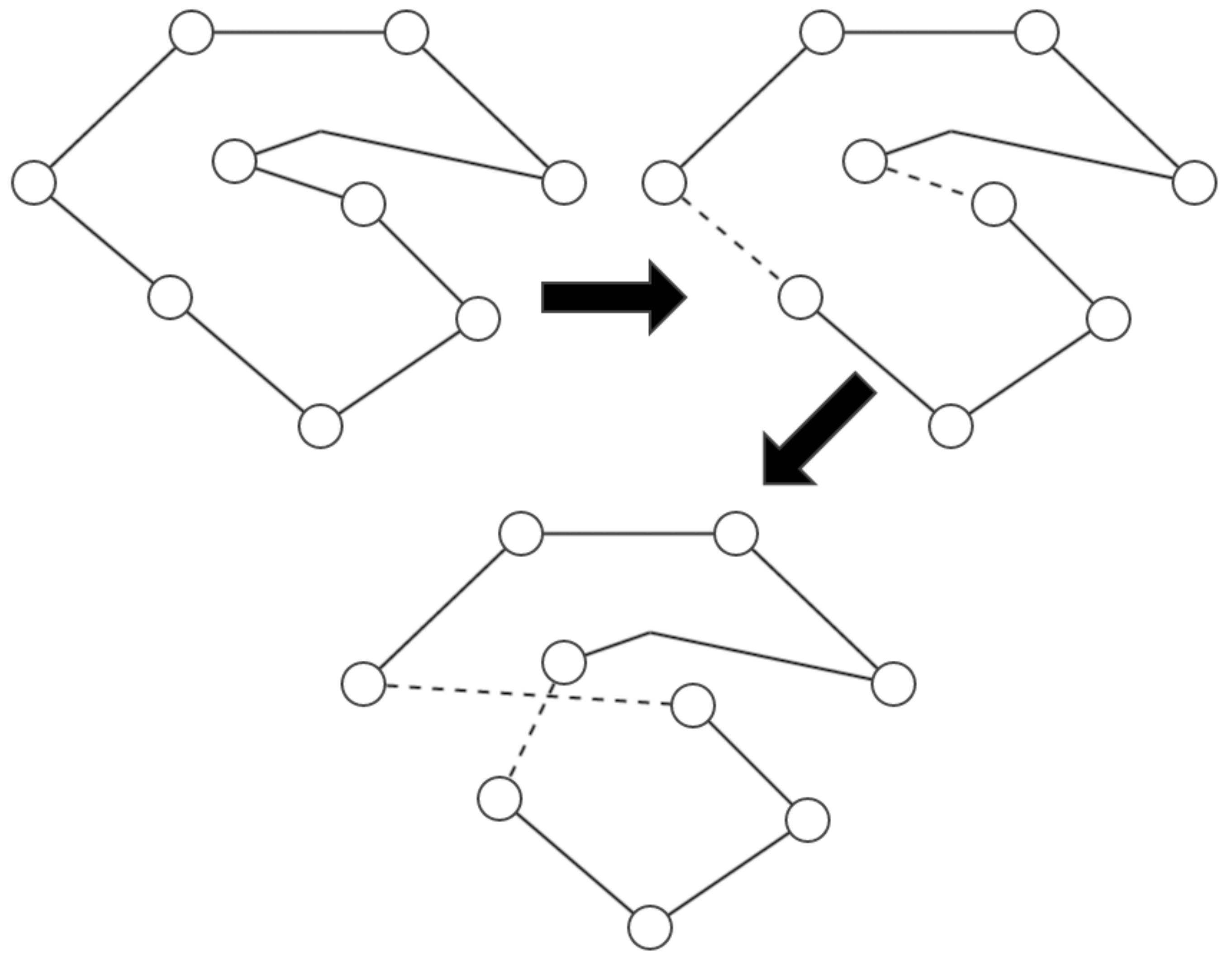
P: Can be solved in polynomial time.

Example: *Long multiplication* algorithm.
Multiplying 2 n-bit integers requires n^2 steps (i.e. $O(n^2)$)

```
        23958233
×           5830
————————————————
        00000000 ( =  23,958,233 ×     0)
       71874699  ( =  23,958,233 ×    30)
     191665864   ( =  23,958,233 ×   800)
+ 119791165      ( =  23,958,233 × 5,000)
————————————————
    139676498390 ( = 139,676,498,390)
```

NP: Can be solved in nondeterministic polynomial time.

Example: *Traveling salesman problem*
Euclidean solution: $O(n(\log n)^{(O(c\sqrt{d}))^{d-1}})$

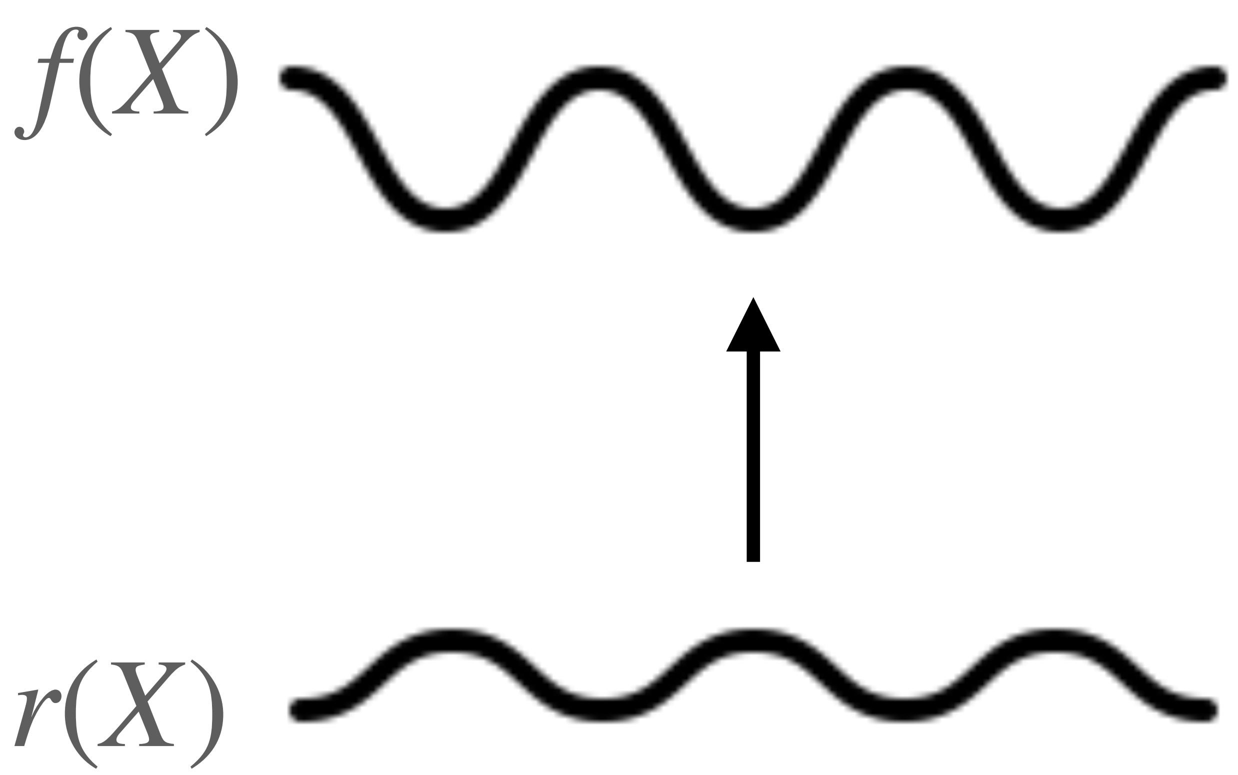


Evolvability

Question: What if life is a structured learning problem?

Process:

1. The goal of cellular life is to produce proteins
2. Functional process:
   - x_i : single DNA segment
   - f : ideal DNA $\rightarrow$ protein distribution
   - r : real DNA $\rightarrow$ protein distribution
   - D_n : distribution over x
3. Goal: Learn the coding (DNA) and translation (RNA) change r to meet the ideal f.



$$\mathrm{Perf}_f(r, D_n) = \sum_{x \in X_n} f(x) r(x) D_n(x)$$

(Valiant 2009)

Structure of a learnable problem

What is learnable?

<u>Learnability theory:</u> Can the true h be learned?

Determine whether the true h is able to be learned given a particular data scenario (e.g., particular $P_{(\mathbf{X},\mathbf{Y})}$)?

<u>Computational theory:</u> Can the solution for the true h be computed?

P: Problem can be *computed* in polynomial time. → *We can only work with "P hard" problems*

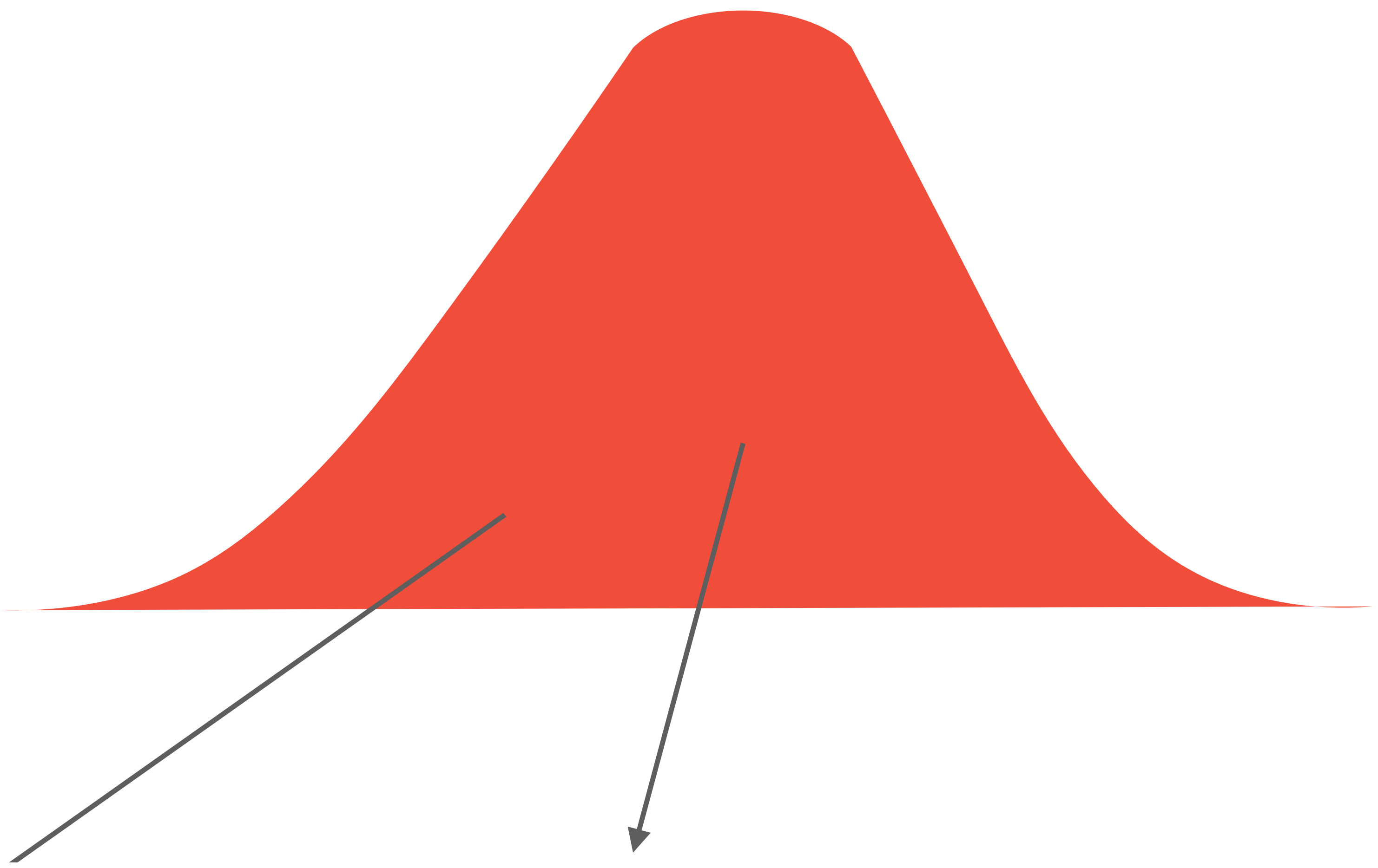
NP: Problem requires non-deterministic polynomial time.

(Fulop & Chater 2013)

Empirical risk minimization

Expected Risk

$$E_{\text{risk}}\left(h, n, P\right) = \underbrace{\int_{(\mathbf{X},\mathbf{Y})_n}}_{\text{train}} \underbrace{R(h)}_{\text{risk}} \underbrace{dP_{(X,Y)_n}}_{\text{distribution}}$$



Assumption: Both the training and test data come from the same distribution.

(Vlastelica 2019; Vapnik 1991)

Empirical risk minimization

$$h_{true} \rightarrow h_{best}$$

<u>Generalization:</u> Given a fixed training data set, find the model that *best* predicts future unseen (test) data set.

$$E_{\text{risk}}\left(h, n, P\right) = \underbrace{\int_{(\mathbf{X},\mathbf{Y})_n}}_{\text{test}} \underbrace{\int_{(\mathbf{X},\mathbf{Y})}}_{\text{training}} \underbrace{\ell(h(X), Y)}_{\text{loss function}} \underbrace{dP_{X,Y}}_{\text{distribution}} \underbrace{dP_{(X,Y)_n}}_{\text{distribution}}$$

(Vlastelica 2019; Vapnik 1991)

Probably Approximately Correct (PAC) Learning

Q: How do you get "good enough" learning so as to be useful?

PAC learning requires a learner to:

1. Approximate the true h
2. Be computationally feasible (P problem)

Approximately: A hypothesis $h \in H$ is approximately correct if its error over the training data $P_{(X,Y)}$ is bounded by some ϵ, with $0 \leq \epsilon \leq \frac{1}{2}$.

Probably: The h is probably approximately correct at a generalization error rate of δ, if its prediction accuracy is $1 - \delta$, with $0 \leq \delta \leq \frac{1}{2}$.

(Valiant 1984; 2009)

Evolvability

Definition: Let F be a class of functions, R be a class of representations (e.g., $r(x)$), and D a class of distributions on X. Then F is *evolvable* by R over D if there exists a set of *polynomial* translation, sample, selection, and tolerance functions such that for all n variables and all $\epsilon > 0$, for all ideal functions $f \in F_n$ and representations $r_0 \in R_n$, with probability at least $1 - \epsilon$

$$Perf(f, r_{g(n,\frac{1}{\epsilon})}) \geq 1 - \epsilon$$

where the sizes of neighbors $N(r)$ for $r \in R_n$ are at most $p(n, \dfrac{1}{\epsilon})$, samples size is $s(n, \dfrac{1}{\epsilon})$, tolerance is $t(\dfrac{1}{n}, \epsilon)$, and generation size is $g(n, \dfrac{1}{\epsilon})$.

(Valiant 2009)

Evolvability (translated)

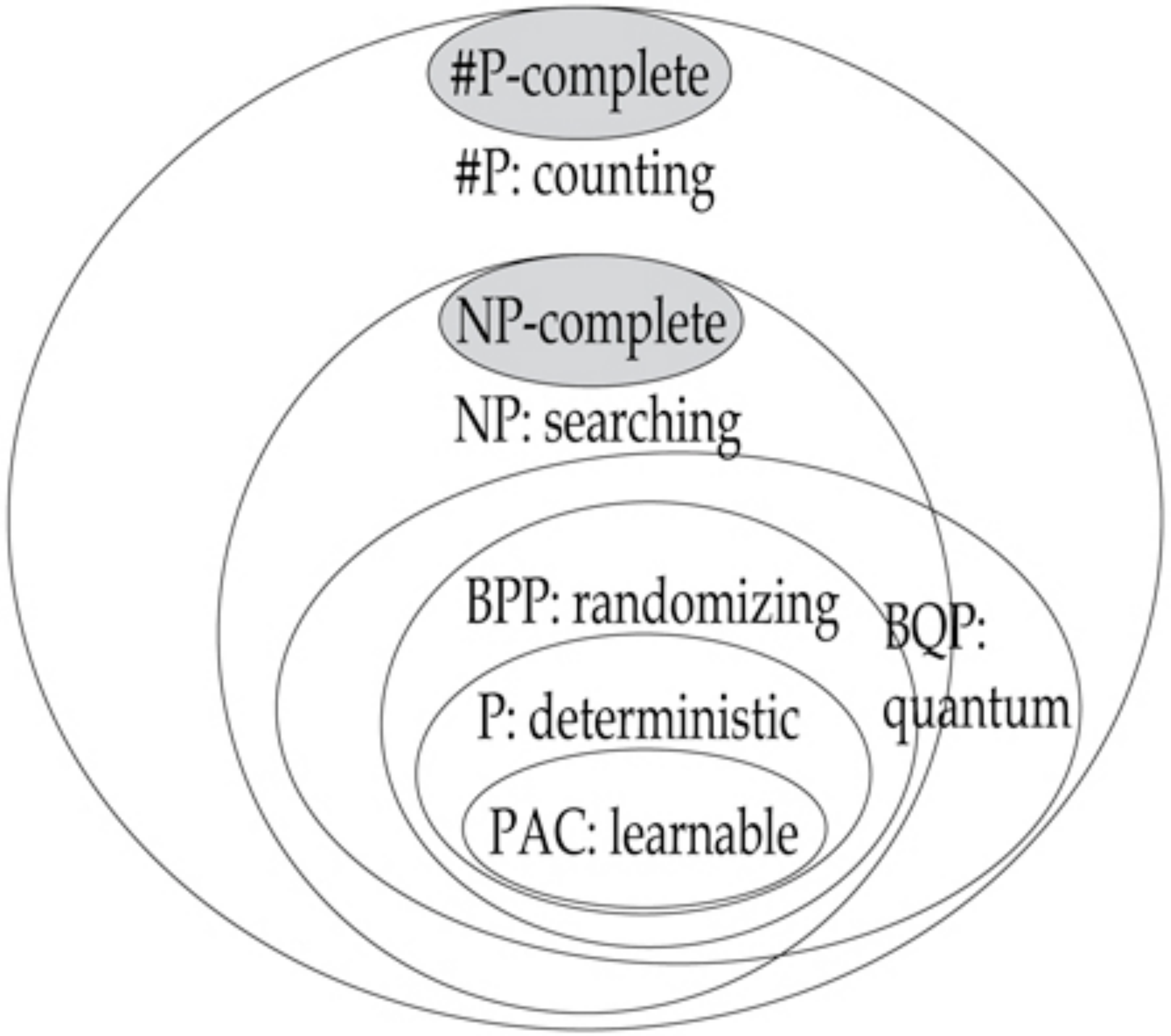
Definition: Assuming a degree of fitness tolerance to error, ϵ, the evolvability of traits can be considered PAC learnable over generations with sufficiently large populations.

Complex life, including intelligent behavior, is possible to be learned *across a reasonable number of generations*. **Thus life can be seen as a learning process.**

(Valiant 2009)

Learnable evolutionary functions

Levels of complexity



Parity function: A Boolean function that has value true if and only if an odd number of its arguments have value true

$$f(x) = x_1 \oplus x_2 \oplus \ldots \oplus x_n$$

"exclusive or"

Complexity: NP-hard

Con(dis)junction Function: A Boolean function that has value true if all of its arguments have value true (or at least one of its arguments are true for disjunctions).

How we think the DNA to protein translation happens

Complexity: P-hard

(Valiant 2009, 2013)

Take home message

- Life as a random search problem is NP-hard and thus unlearnable.
- Life as a structured learning problem is P-hard and learnable in a reasonable (and quick) amount of time.
- Thus, life (including intelligent behavior) is a learning process.

Discussion time

Task: How do you reconcile Schrödinger's two principles of life

- Life controls/reduces entropy of its environment.
- Manipulates local entropy so as to maintain molecular stability.

with Valiant's evolvability argument? Are these saying the same thing or are they different views on the goals of evolution?