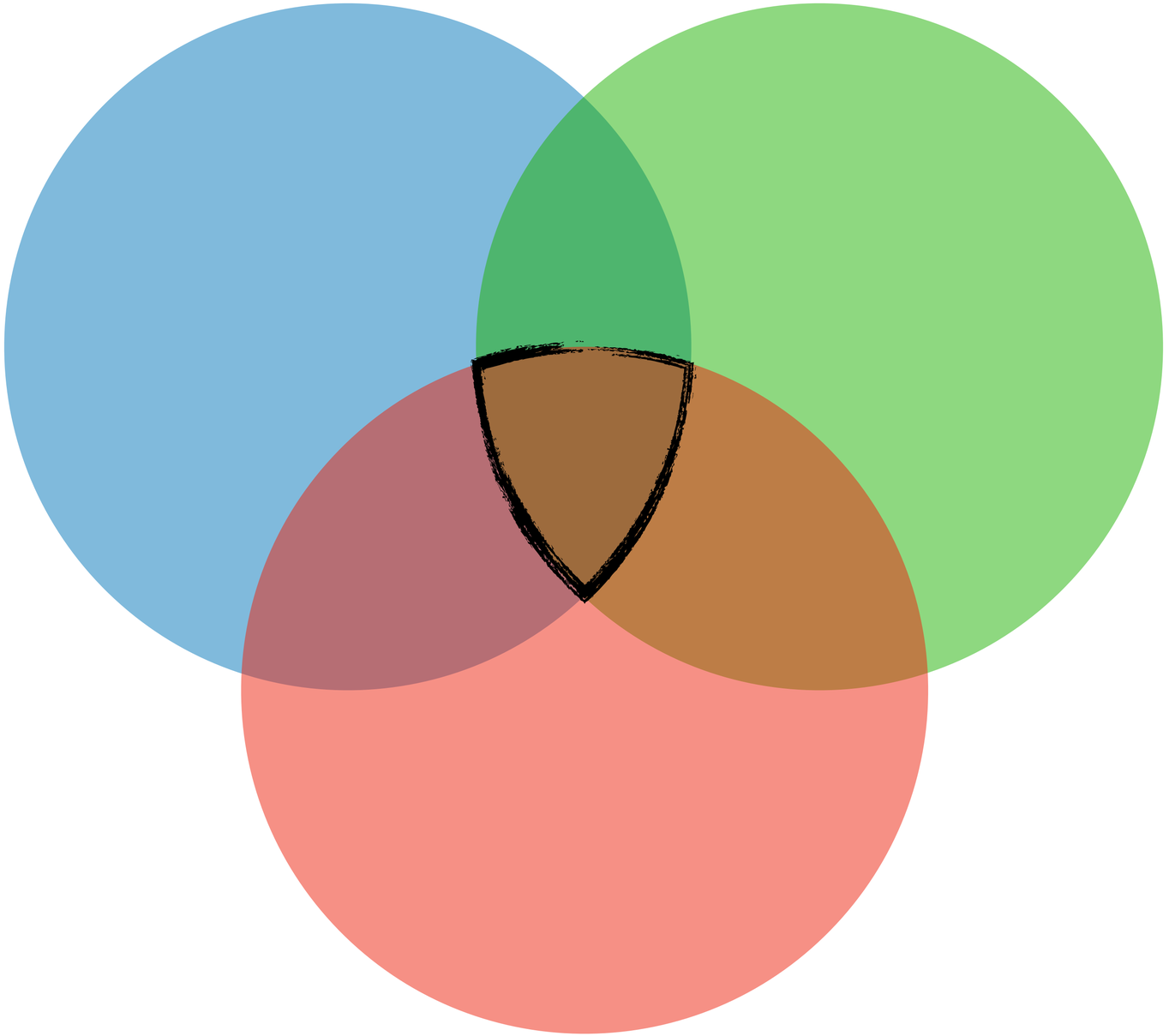
Description and Complexity in the age of Deep Learning

Peter Bloem, Learning and Reasoning, Vrije Universiteit Amsterdam

LEARNING
***and* REASONING**

Minimum Description Length



Kolmogorov Complexity and AIT

Deep Learning

Kolmogorov complexity

An object's Kolmogorov complexity is the length of its smallest practical description.

Algorithmic information theory (AIT): an information theory built on KC.

Minimum description length (MDL): AIT's more practical cousin.

Kolmogorov complexity

in bits

An object's Kolmogorov complexity is the length of its smallest practical description.

Algorithmic information theory (AIT): an information theory built on KC.

Minimum description length (MDL): AIT's more practical cousin.

Kolmogorov complexity

in bits

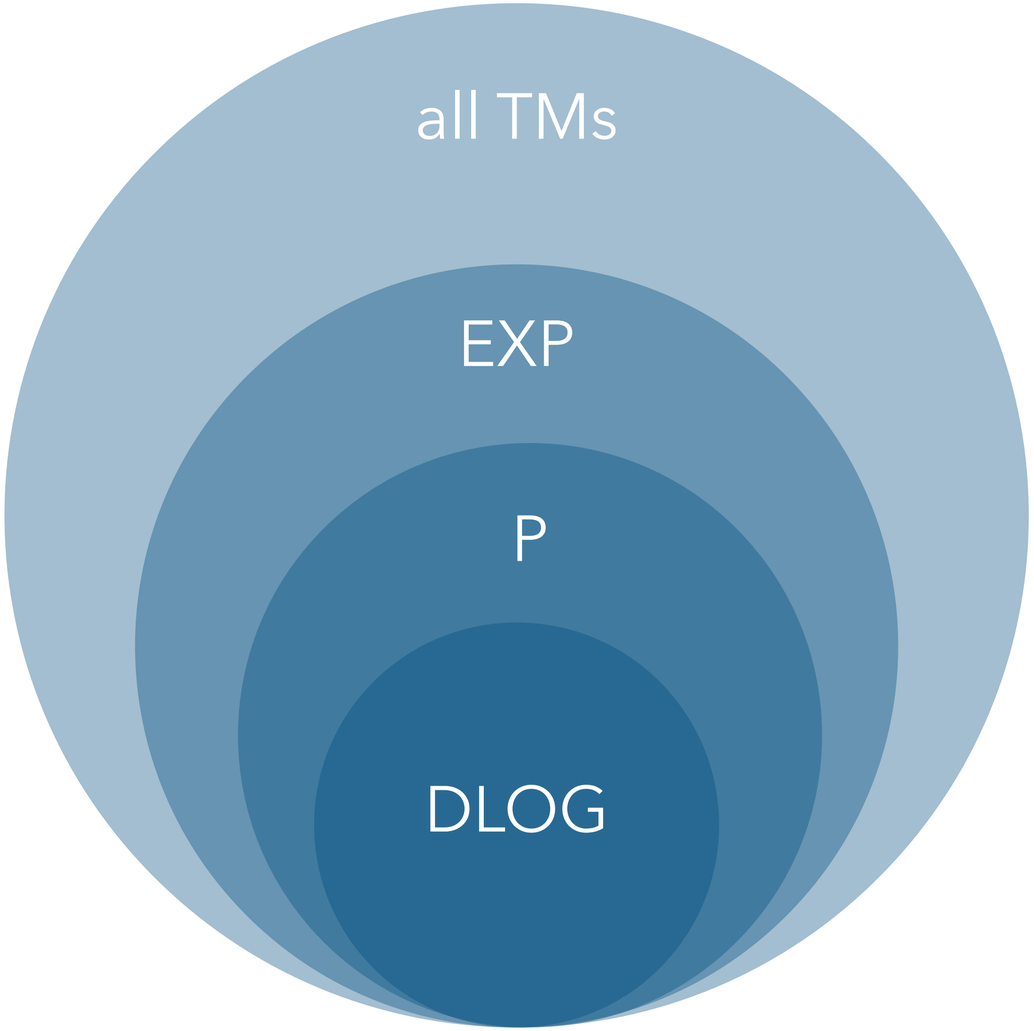
An object's Kolmogorov complexity is the length of its smallest practical description.

on a universal computer

Algorithmic information theory (AIT): an information theory built on KC.

Minimum description length (MDL): AIT's more practical cousin.

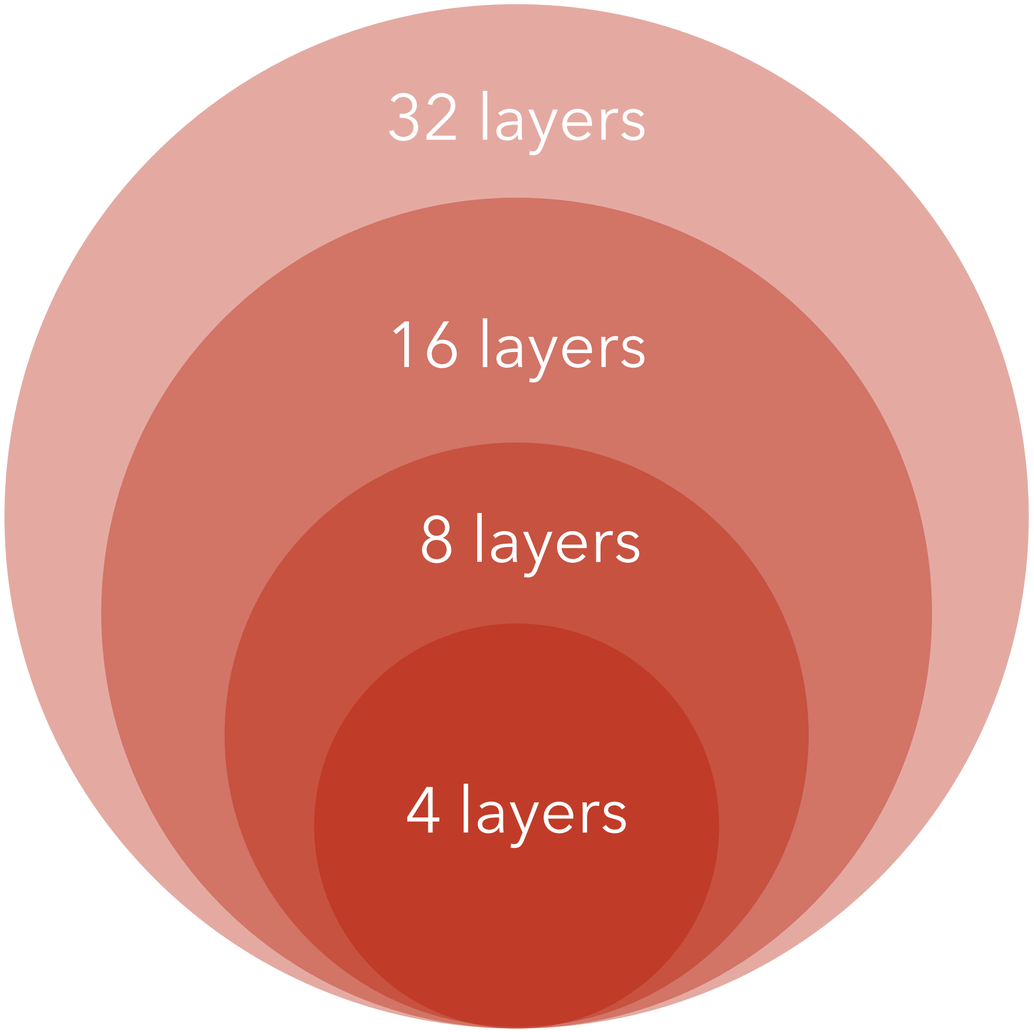
example: resource-bounded complexity



The smallest description that can be unpacked by a TM in the given class.

- How rare are strings that require, say EXP, to reach maximum KC?
- How difficult is it to generate such a string?
- What do these strings look like?

Enter deep learning



Deep learning gives us practical model classes with hierarchical complexity.

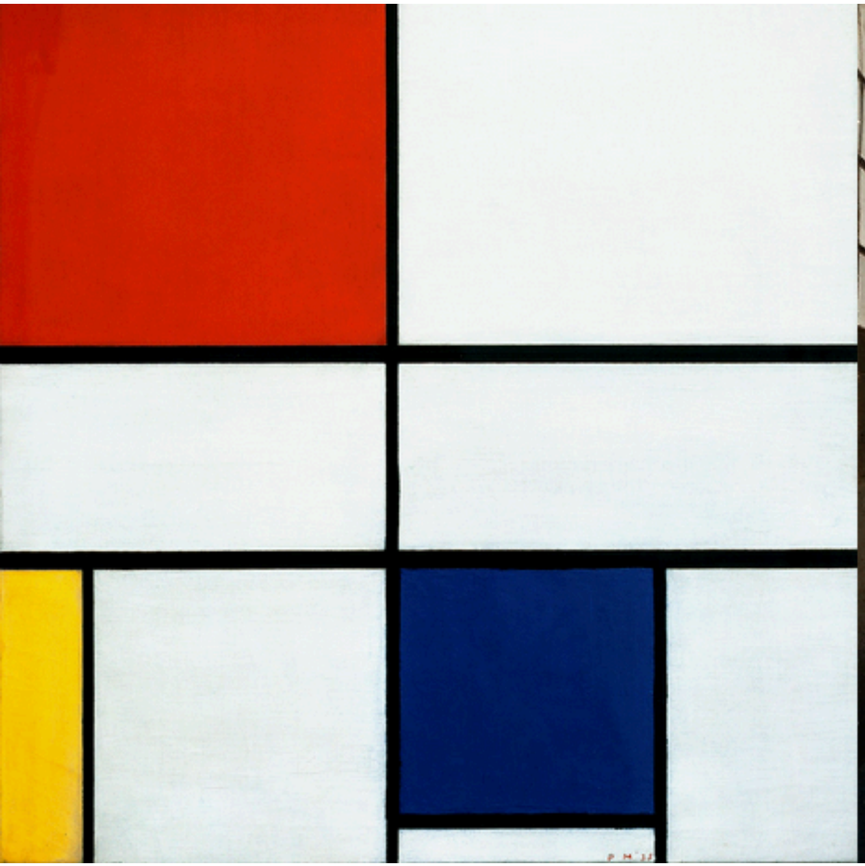
- We didn't have these when most of MDL and AIT was developed.

example 1: safe approximation

If your string x is generated by a model in class C, then $K_C(x) = K(x)$ to within a few bits with high probability (Bloem et al. 2015).
More precisely: to within k bits with a probability that vanishes exponentially in k.

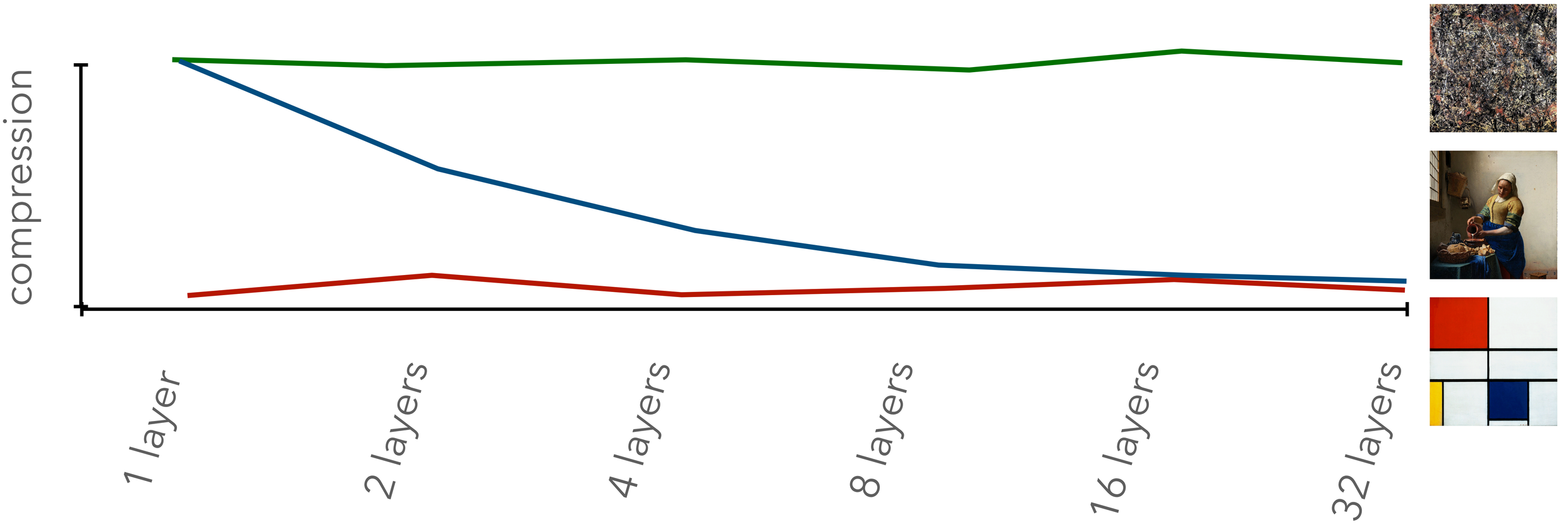
- Safe approximations are *transitive*.
- Does this carry over to NNs? Is the solution found by GD a *safe approximation* of the optimal solution?
- A simple language of *provable optimality*.

example 2: sophistication



neurological depth

logical/computational depth: the amount of computation required to achieve the optimal compression



example 3: mutual information

- Mutual information between distributions: well defined, often computable, estimable.
- Mutual information between *objects*?

Why _?

- Why this? AIT & MDL can answer many pressing questions in modern machine learning.
- Why now? Deep Learning offers, for the first time ever, practical models that are *hierachical* in complexity.
- Why us? Both MDL and AIT were developed in no small part in Amsterdam.

What else? Complexity theory. Statistics. Learning theory. Reinforcement learning. Inductive biases.