

Scaling-up AI Systems: Insights from Computational Complexity

Bart Selman & Carla Gomes
Cornell University

Ryan Williams
Stanford → MIT



Computational Complexity Hierarchy

EXP-complete:

games like Go, ...

PSPACE-complete:

QBF, *planning*, chess
(bounded), ...

#P-complete/hard:

#SAT, sampling,
probabilistic inference, ...

NP-complete:

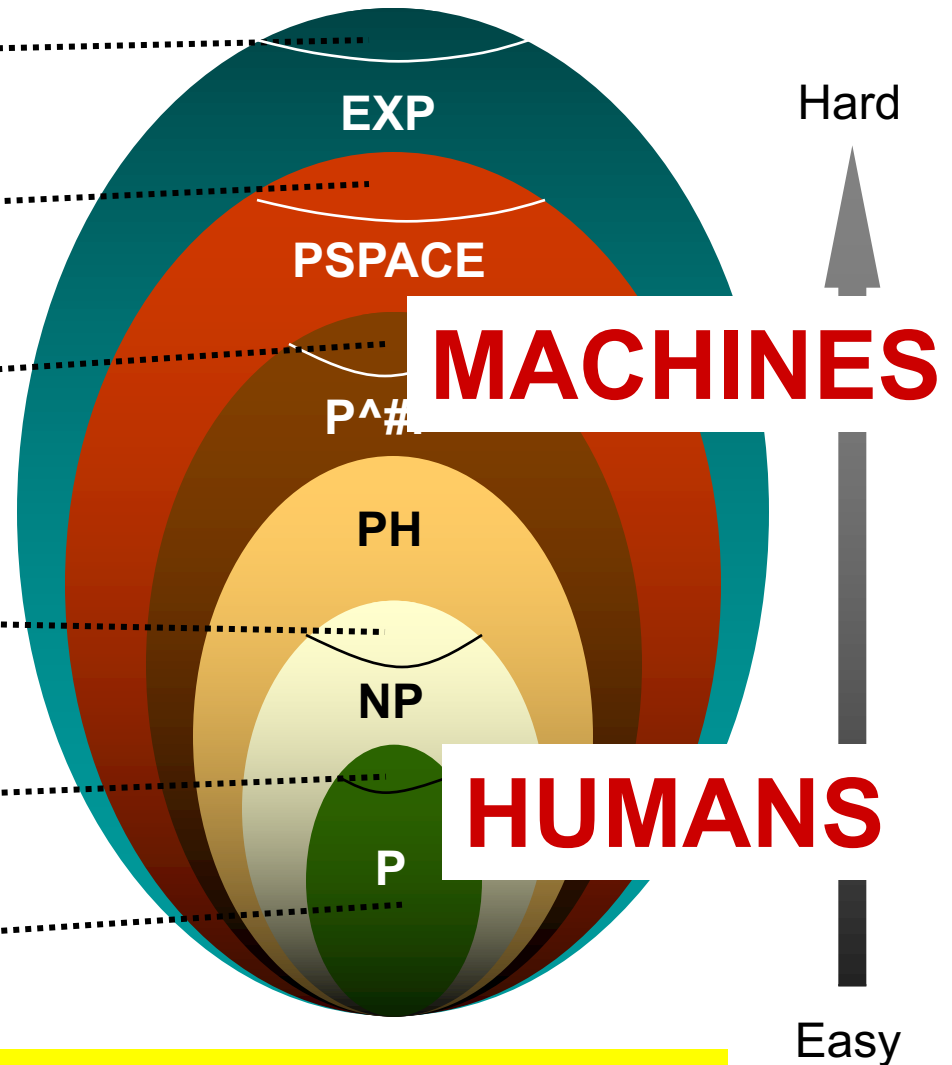
SAT, deep learning, propositional
reasoning, scheduling ...

P-complete:

circuit-value, ...

In P:

DB, sorting, shortest path, ...



What are the implications for human understanding of machine intelligence?

Focus: Human understanding of super-intelligent machines

Hypothesis: Even though machines are moving to higher levels of the computational complexity hierarchy, it may not necessarily be the case that humans won't be able to understand their behaviors/decisions.

Why? In earlier work, we showed how automated reasoning on very large reasoning problems (millions of variables) can often be understood in terms of the behavior of a small set (a few dozen) of key variables (“backdoor variables”). The machine can provide the backdoor variables (i.e., **explains itself**).

$B(n)$	deterministic	randomized	heuristic
n/k	small $exp(n)$	smaller $exp(n)$	tiny $exp(n)$
$O(\log n)$	$\left(\frac{n}{\sqrt{\log n}}\right)^{O(\log n)}$	$\left(\frac{n}{\log n}\right)^{O(\log n)}$	$poly(n)$
$O(1)$	$poly(n)$	$poly(n)$	$poly(n)$

Fc - + + - + - - + + - + + - + - - + - - + + - + - - + - - +
su + - + - - + + + - - + - - + + - + - - + + - + + - - + -
 + - + + - + - - + + - + - - + - - + - + + - + + - + + -
 + + - + - - + - - + + - + + - + - - + + - + - + + - + + -
 - + - - - + + + - - + - - + - - + + + - + + - + + - + + -
 - - + - - + + + - - + - - + - - + + + - + + - + + - + + -
 + - - + + - + - - + + - + + - + - - + - - + - - + - - + -
 - - + + + - + - - + + - + + - + - - + - - + - - + - - + -
 + - + - - + + - - + + - + + - + - - + - - + - - + - - + -
 + + - - + - - + + - - + + - + + - + - - + - - + - - + -
 - + + - + - - + + - - + - - + + - - + + - - + + - - + + -
 - - + + - + - - + + - - + - - + + - - + + - - + + - - + + -
 + - - + - - + + + - - + - - + + + - - + - - + + - - + - - +
 + - + + - - + + - - - + + - + - - + + - - + + - - + - - +
 - + - - + + - - + - - + + - - + + - - + + - - + + - - + + -
 + + + - + - - - + + - - + - + + - + - + + - - + + - + - -
 - + + - + - + + - - + + - + - - + - + + - + - - + + - -
 - - - + - + - + + - - + + - + + - + + - + + - + + - + - -
 + - - + - - + + + - - + + - - - + - - + - - + - - + - - + +
 + - - + - + + - - + - - + - - + - - + - - + - - + - - + - +



So, at least in
for human un

1160 --- +1 / -1s elements

all sums of sub-sequences

stay between

-2 and +2

No 1161 sequence exists!

More formall
computational

Concrete chal
discrepancy co
human access
(likely... we c

ere is hope
ning.

r complex
typical case)

he *Erdos*
2014). Does a

