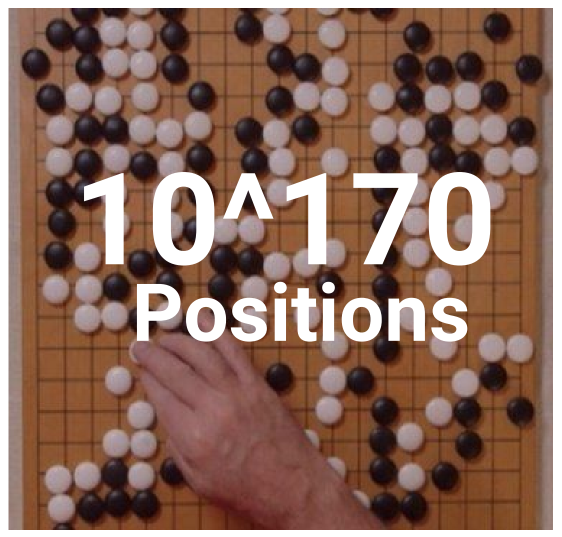
AlphaGo

Go in numbers

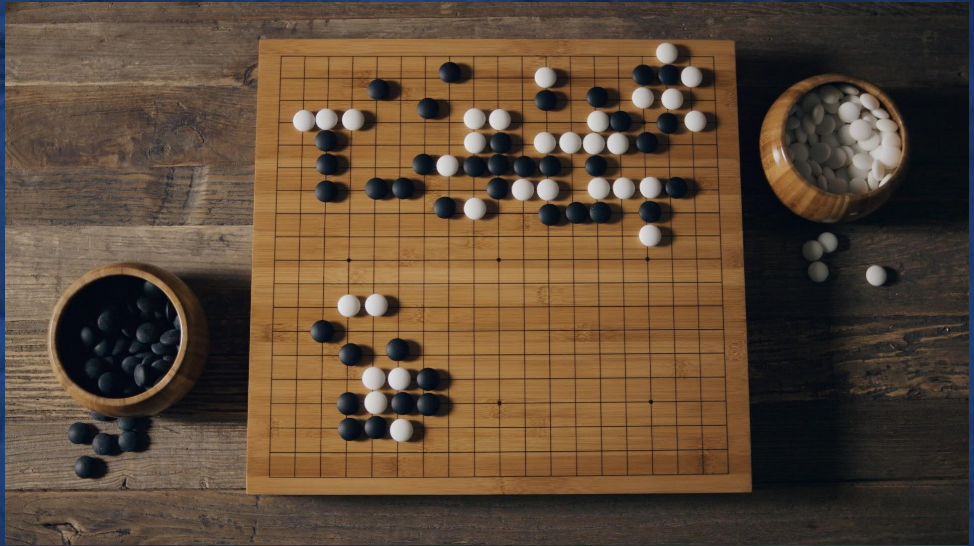


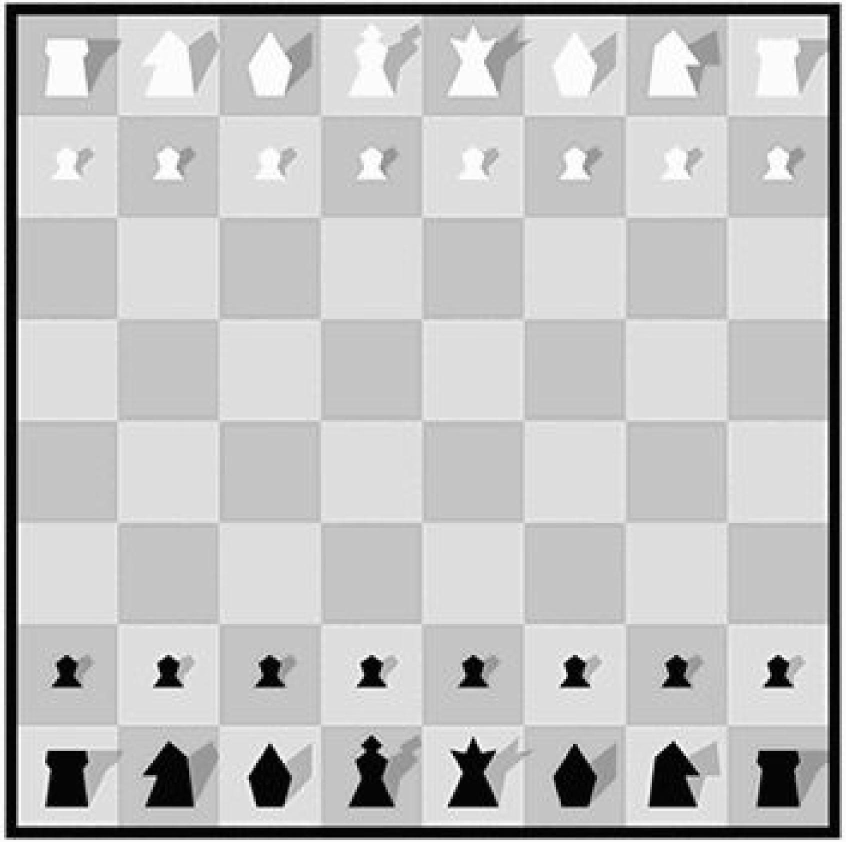
Why is Go hard for computers to play?

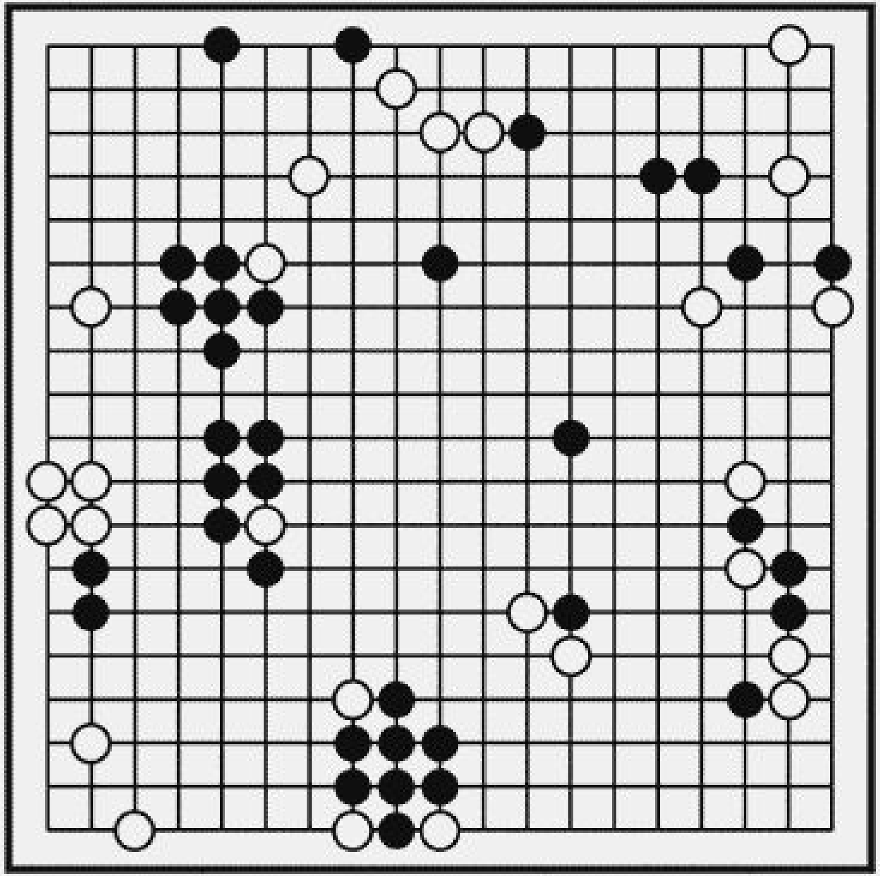
Game tree complexity = b^d

Brute force search intractable:

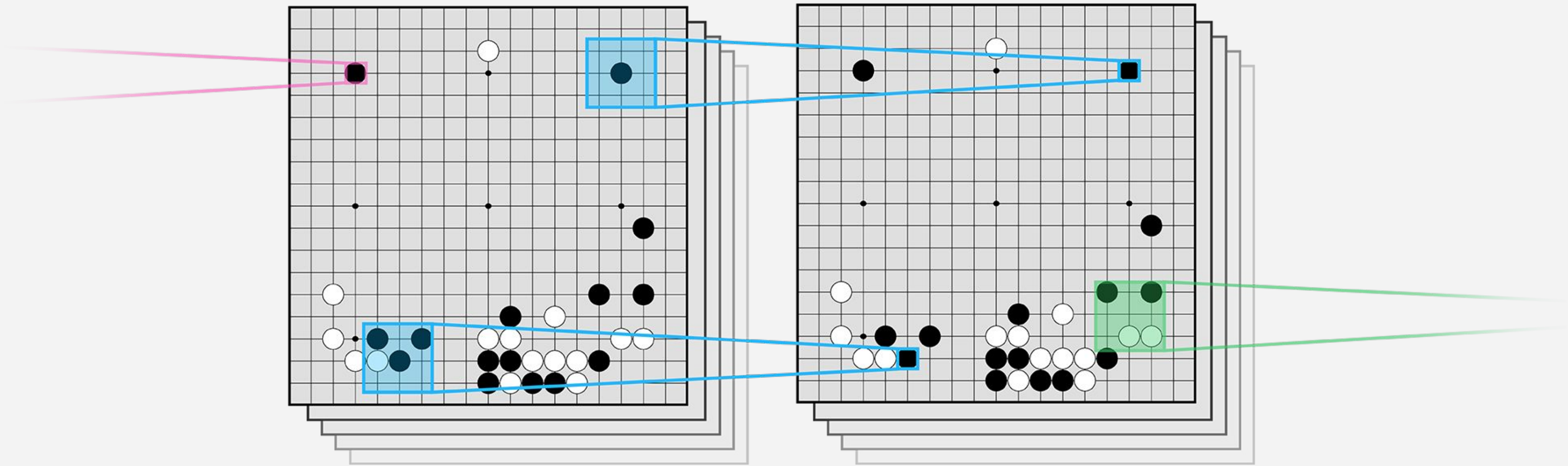
1. Search space is huge
2. "Impossible" for computers to evaluate who is winning



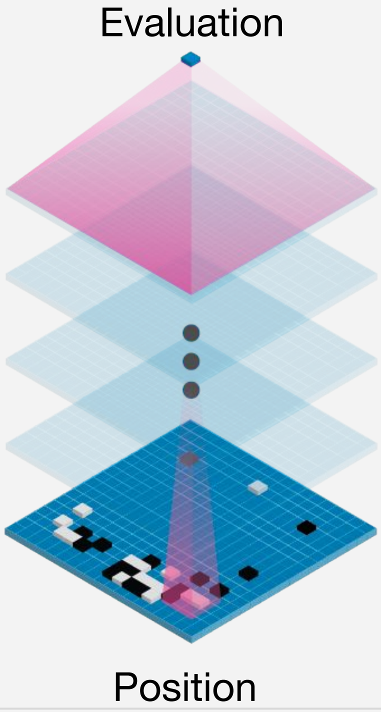


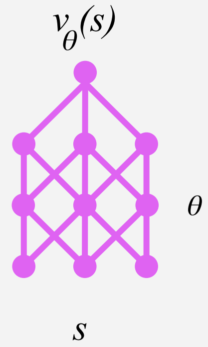


Convolutional neural network

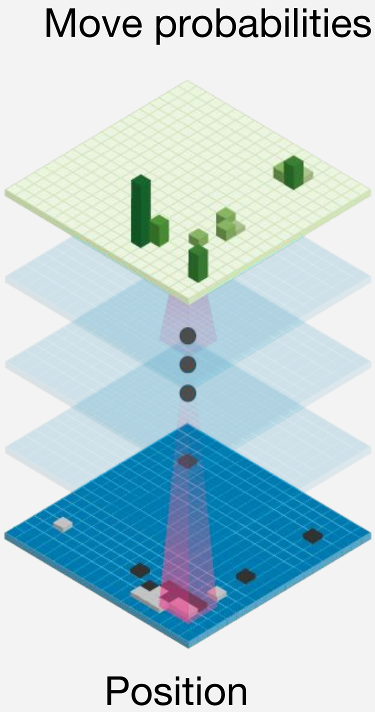


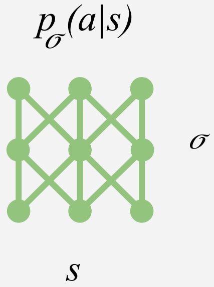
Value network



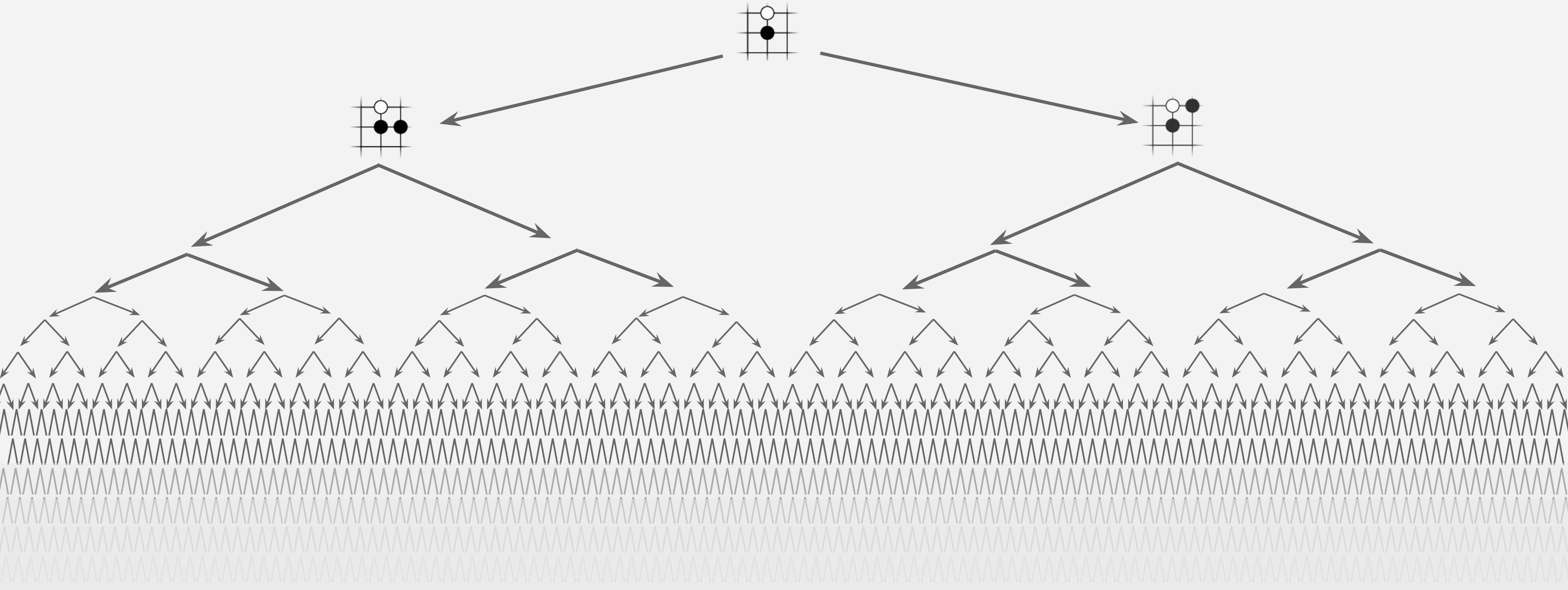


Policy network

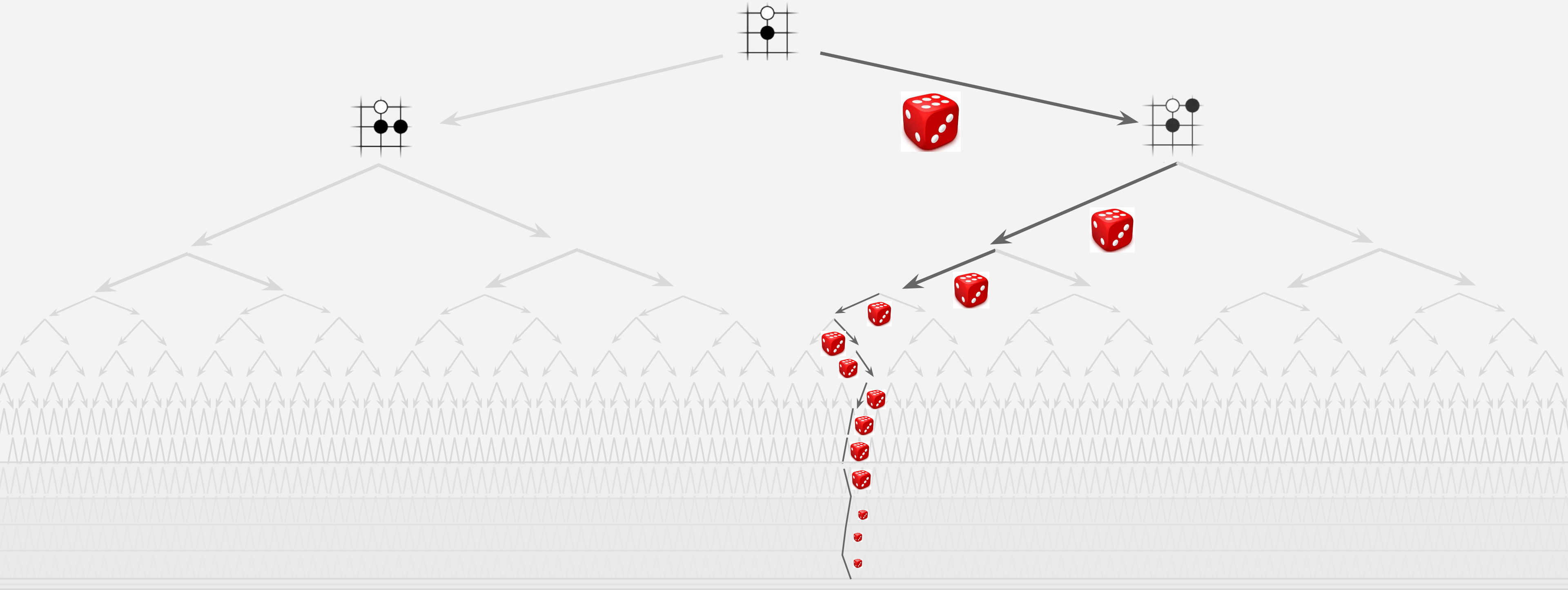




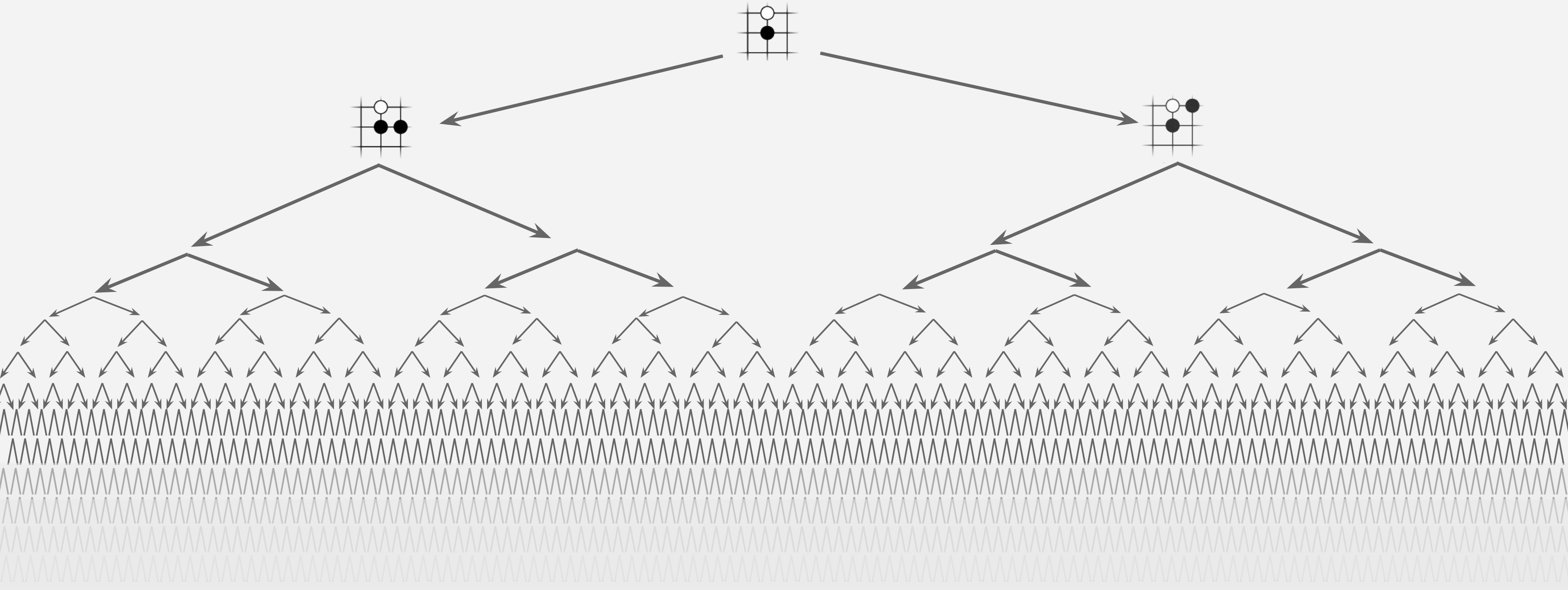
Exhaustive search



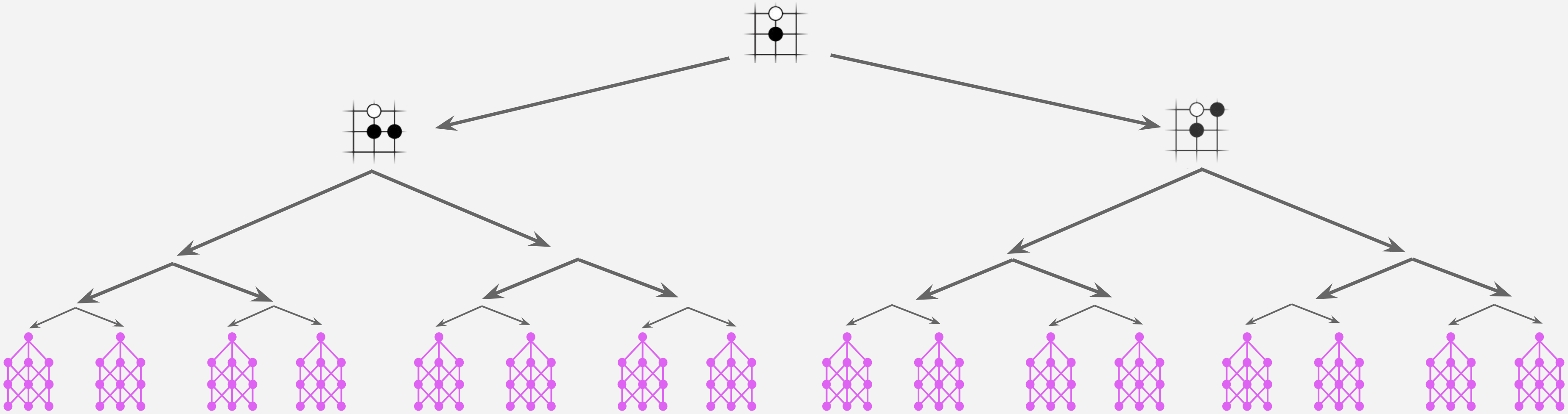
Monte-Carlo rollouts



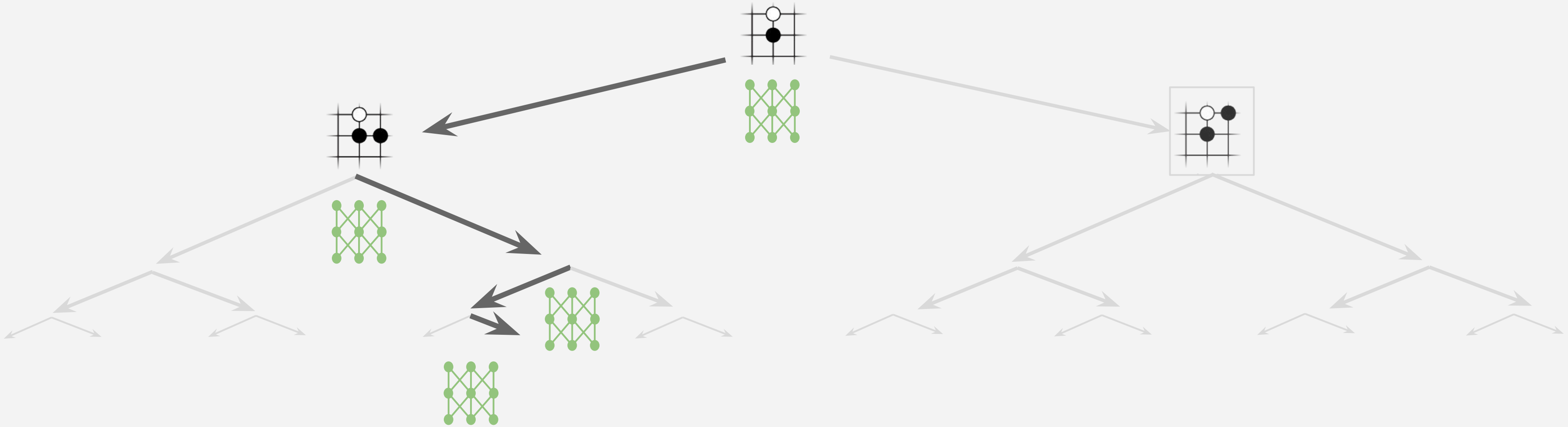
Reducing depth with value network



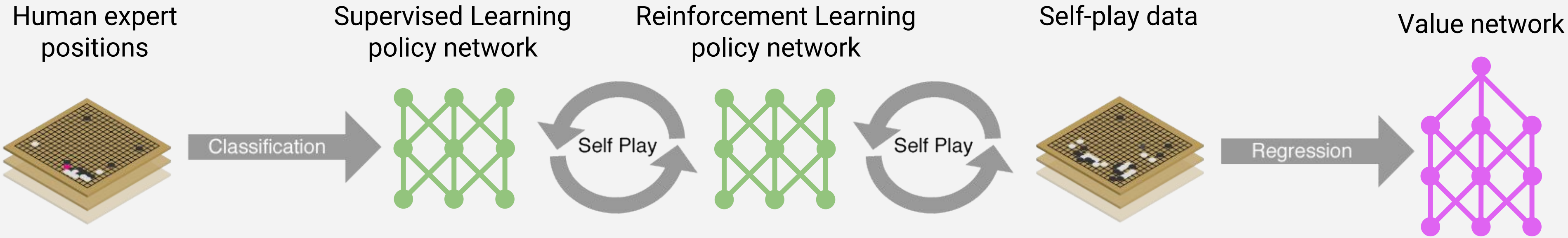
Reducing depth with value network



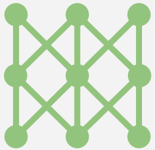
Reducing breadth with policy network



Neural network training pipeline



Supervised learning of policy networks



Policy network: 12 layer convolutional neural network

Training data: 30M positions from human expert games (KGS 5+ dan)

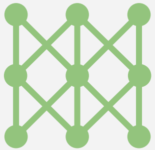
Training algorithm: maximise likelihood by stochastic gradient descent

$$\Delta\sigma \propto \frac{\partial \log p_\sigma(a|s)}{\partial \sigma}$$

Training time: 4 weeks on 50 GPUs using Google Cloud

Results: 57% accuracy on held out test data (state-of-the art was 44%)

Reinforcement learning of policy networks



Policy network: 12 layer convolutional neural network

Training data: games of self-play between policy network

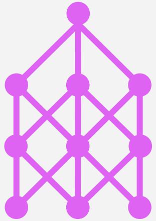
Training algorithm: maximise wins z by policy gradient reinforcement learning

$$\Delta\sigma \propto \frac{\partial \log p_\sigma(a|s)}{\partial \sigma} z$$

Training time: 1 week on 50 GPUs using Google Cloud

Results: 80% vs supervised learning. Raw network ~3 amateur dan.

Reinforcement learning of value networks



Value network: 12 layer convolutional neural network

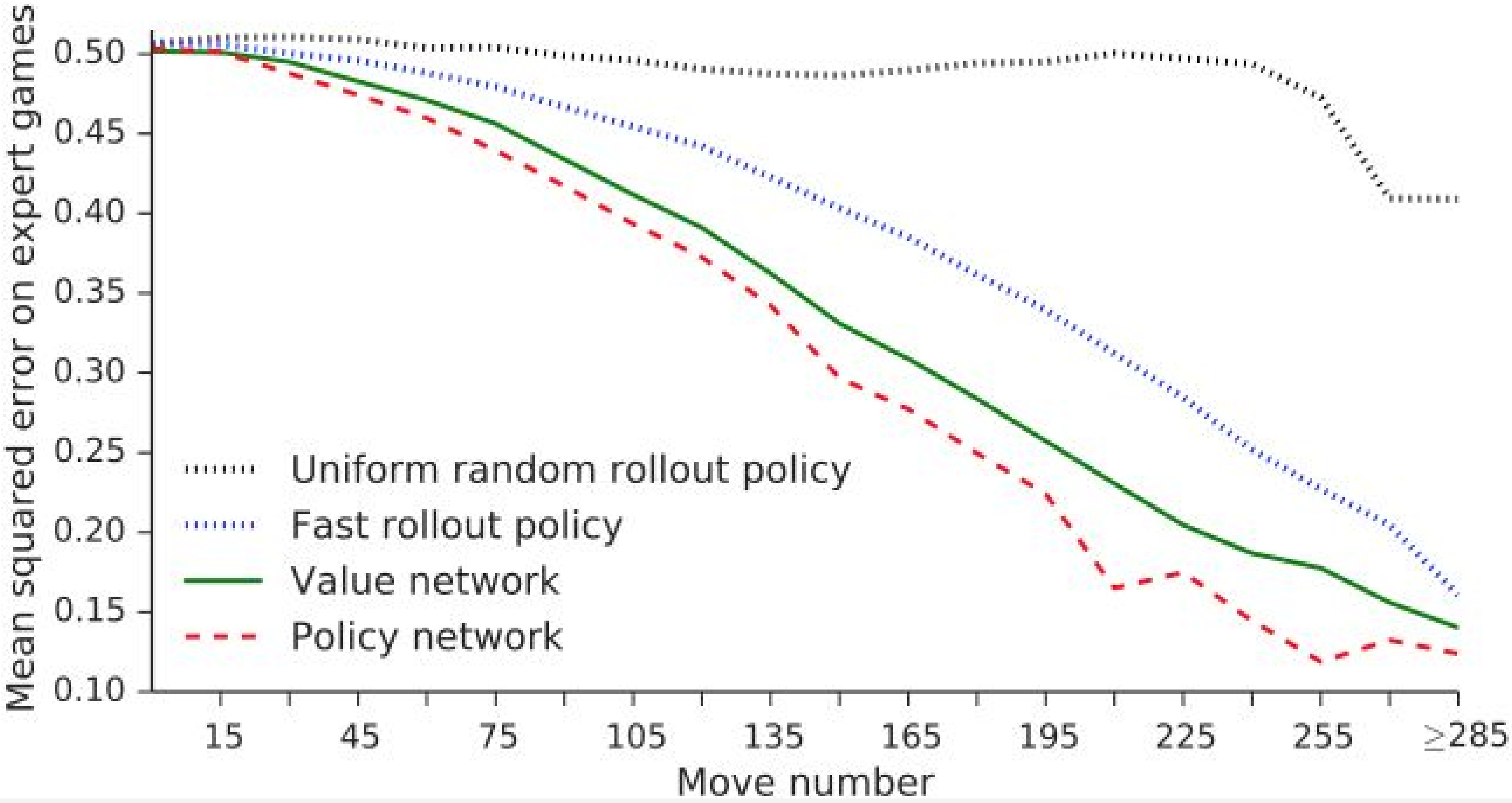
Training data: 30 million games of self-play

Training algorithm: minimise MSE by stochastic gradient descent

$$\Delta\theta \propto \frac{\partial v_\theta(s)}{\partial\theta}(z - v_\theta(s))$$

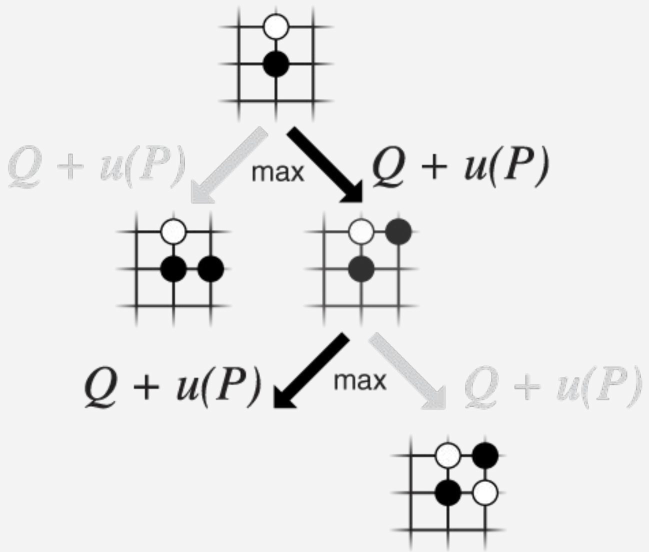
Training time: 1 week on 50 GPUs using Google Cloud

Results: First strong position evaluation function - previously thought impossible



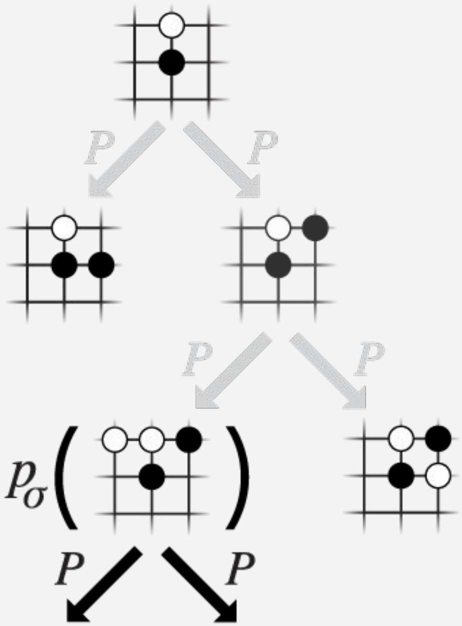
Mean squared error on expert games
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
15
45
75
105
135
165
195
225
255
≥285
Move number
Uniform random rollout policy
Fast rollout policy
Value network
Policy network

Monte-Carlo tree search in AlphaGo: **selection**



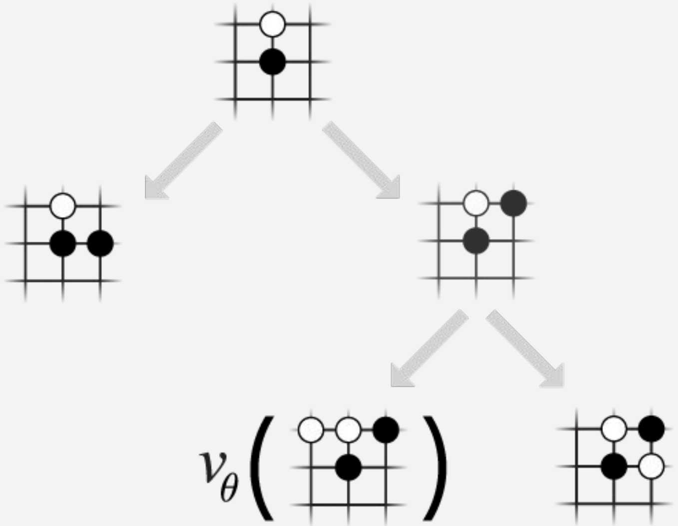
P prior probability
Q action value

Monte-Carlo tree search in AlphaGo: **expansion**



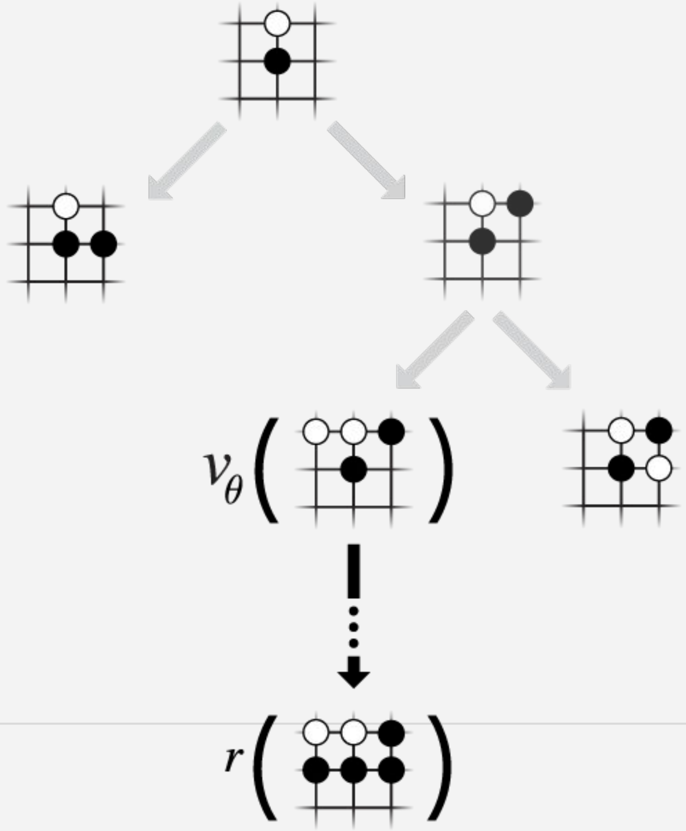
p_σ Policy network
P prior probability

Monte-Carlo tree search in AlphaGo: **evaluation**



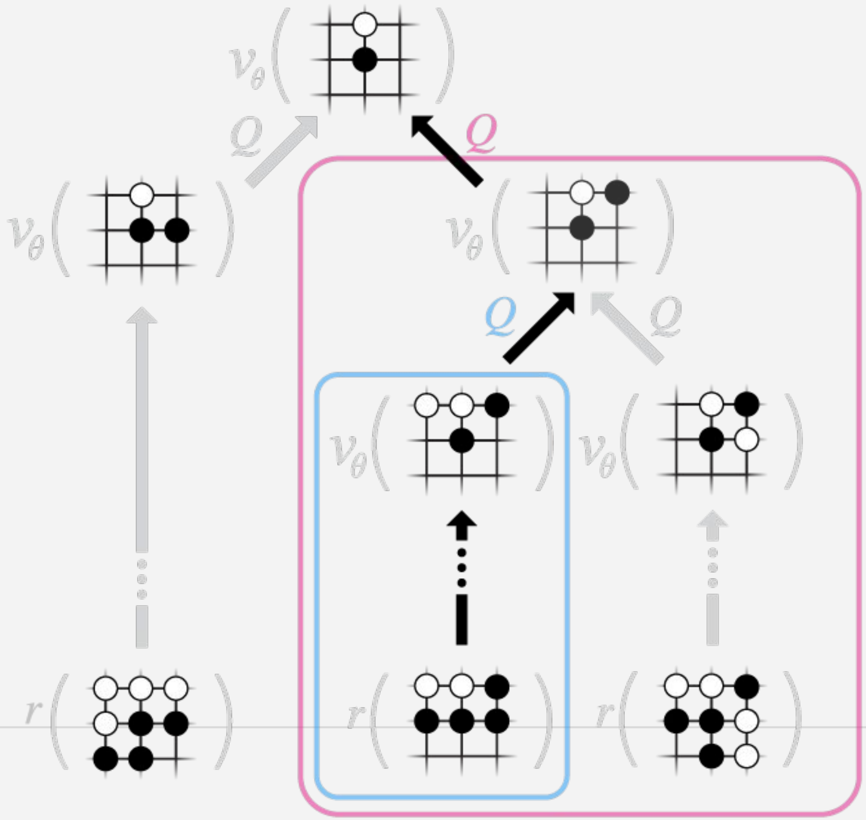
v_θ Value network

Monte-Carlo tree search in AlphaGo: **rollout**



v_θ Value network
r Game scorer

Monte-Carlo tree search in AlphaGo: **backup**



Q Action value
v_θ Value network
r Game scorer

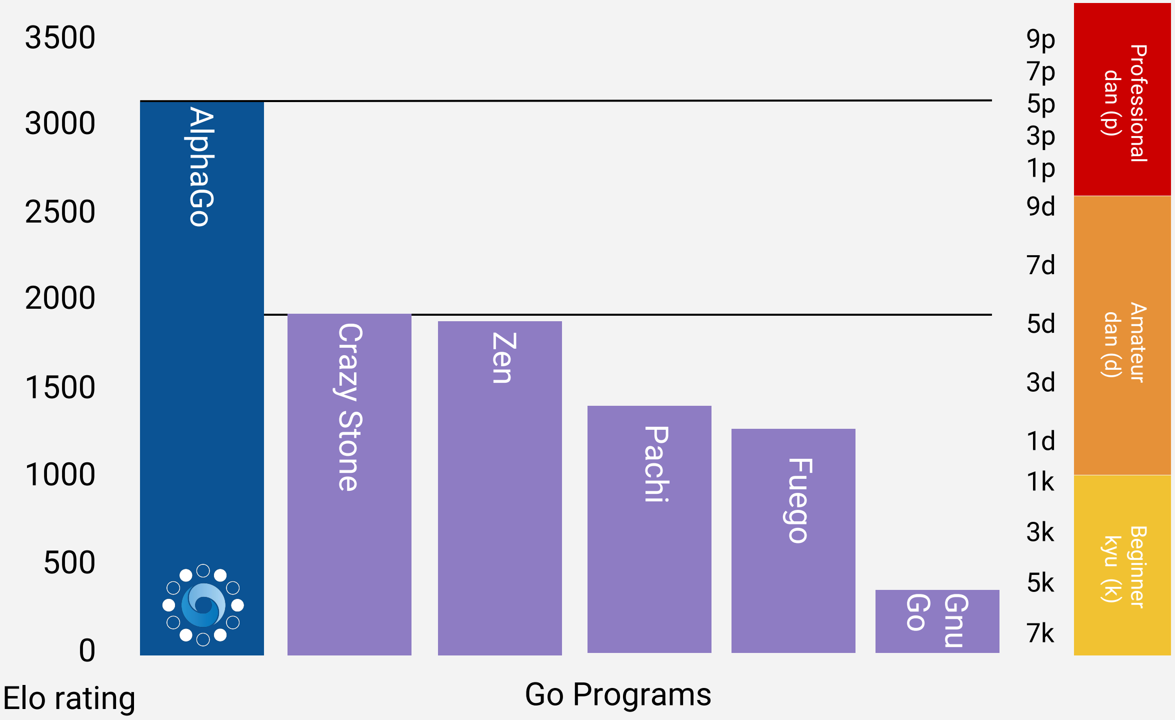

nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card 40 years on
PAGE 462
NATUREASIA.COM

Evaluating Nature AlphaGo against computers

494/495 against computer opponents

>75% winning rate with 4 stone handicap

Even stronger using distributed machines



Evaluating Nature AlphaGo against humans

Fan Hui (2p): European Champion 2013 - 2016

Match was played in October 2015

AlphaGo won the match 5-0

First program ever to beat a professional

on a full size 19x19 in an even game

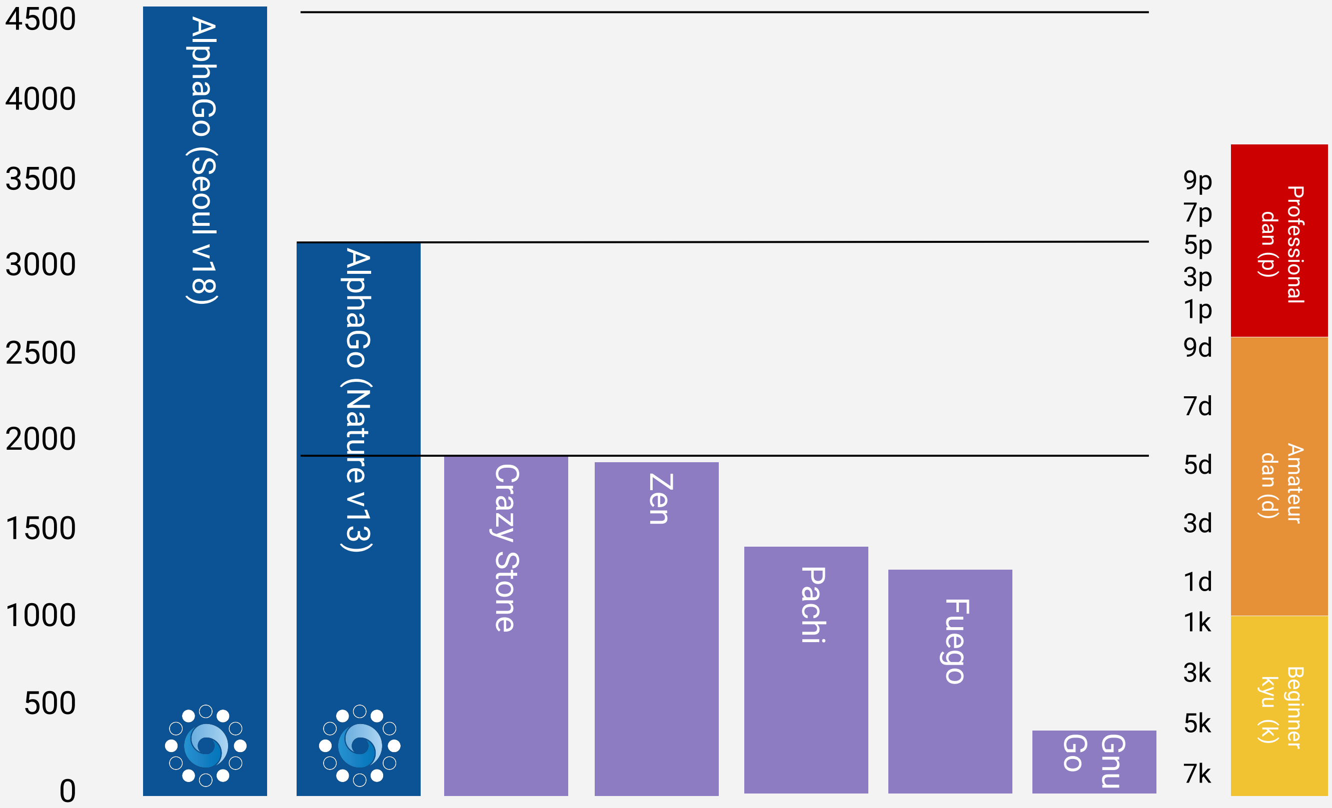
Seoul AlphaGo: Improvements

- Improved value network
- Improved policy network
- Improved search
- Improved hardware (TPU vs GPU)

Evaluating Seoul AlphaGo against computers

Beats Nature AlphaGo with 3 to 4 stones handicap

CAUTION: ratings based on self-play results

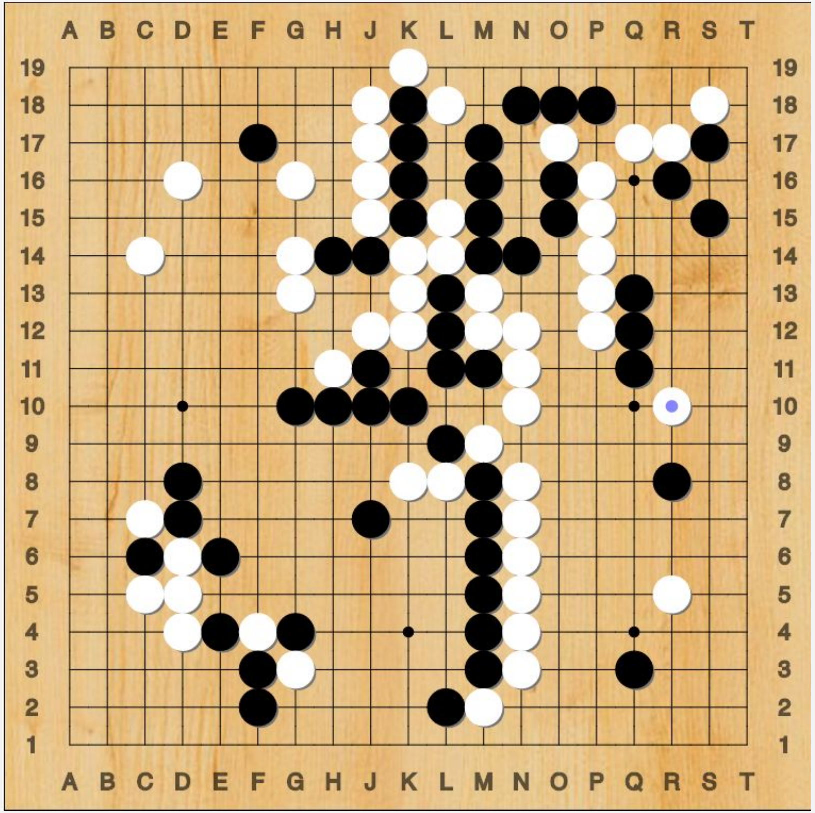


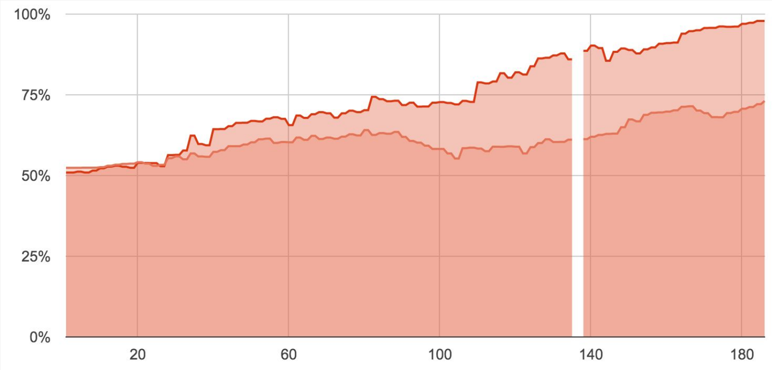

KBA
Korea Baduk Association


KBA
Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo
AlphaGo
Lee Sedol
Google DeepMind
Challenge Match

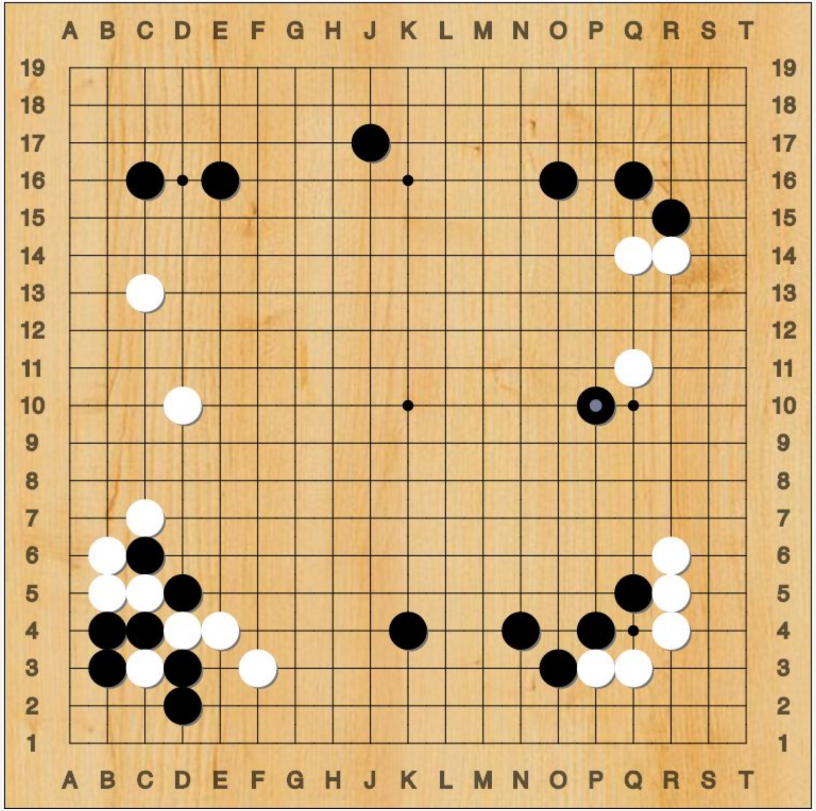

KBA
Google DeepMind
Challenge Match
8 - 15 March 2016

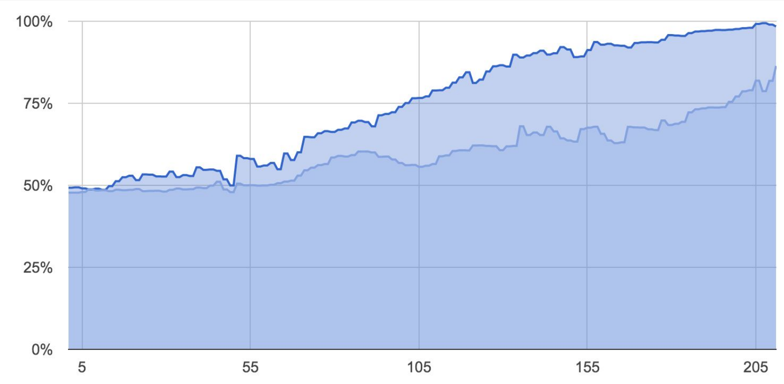
AlphaGo vs Lee Sedol: Game 1



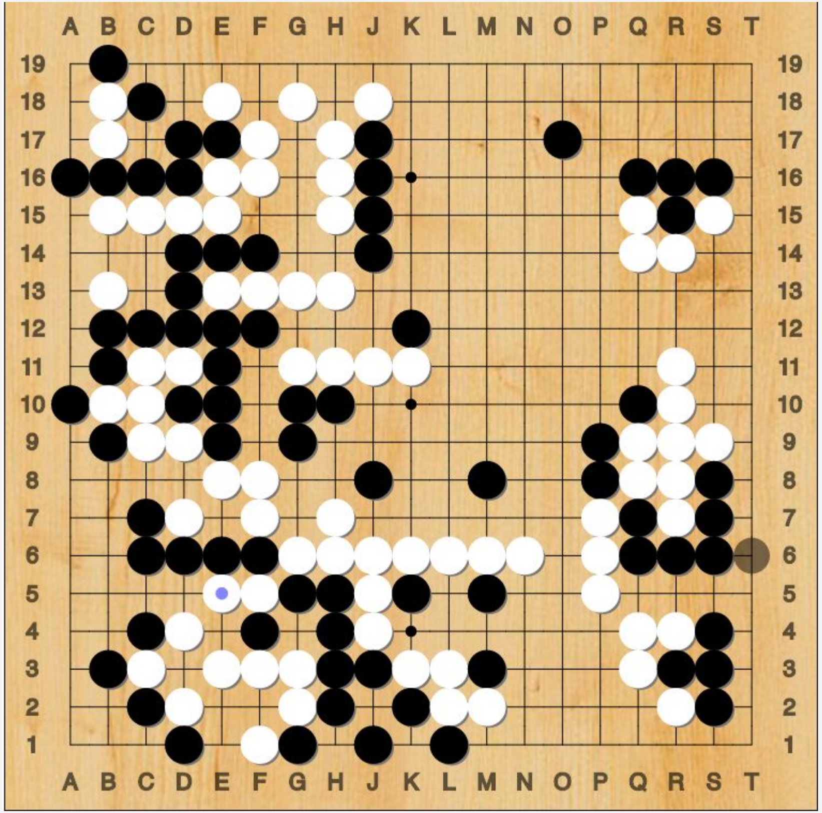


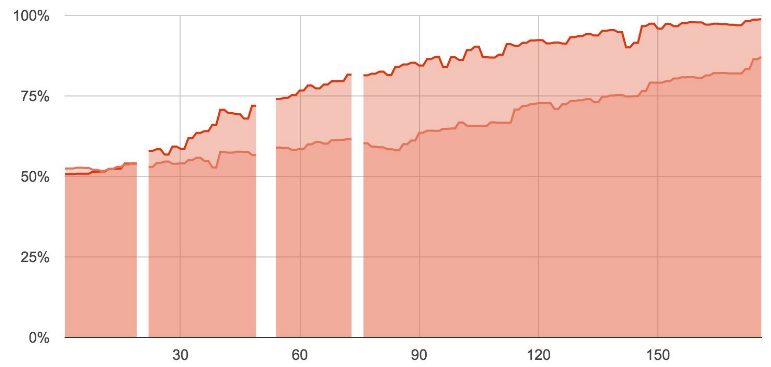
AlphaGo vs Lee Sedol: Game 2



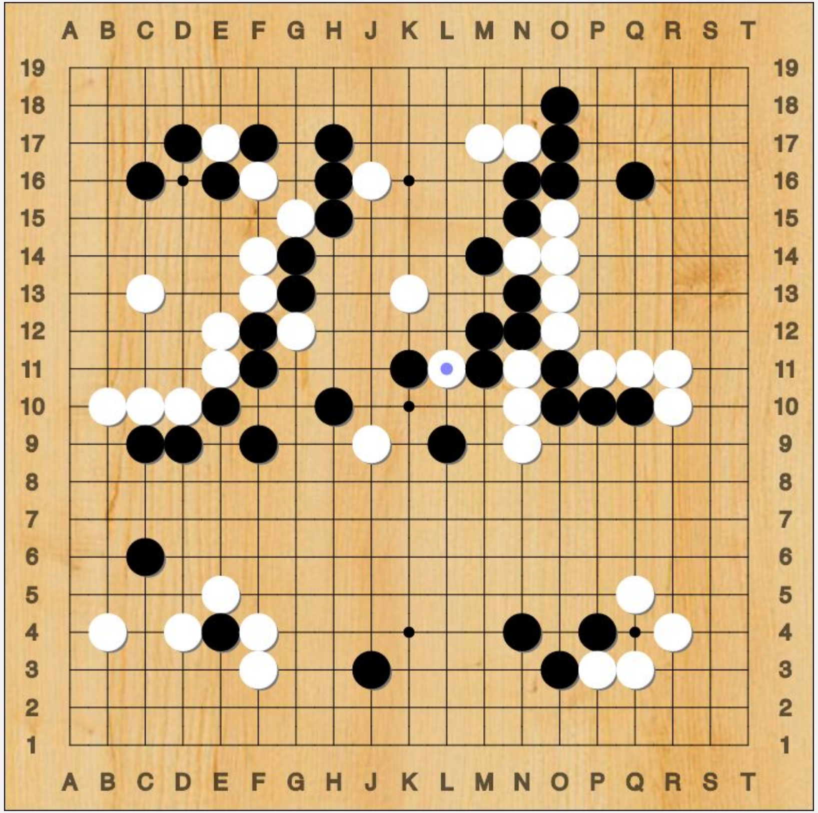


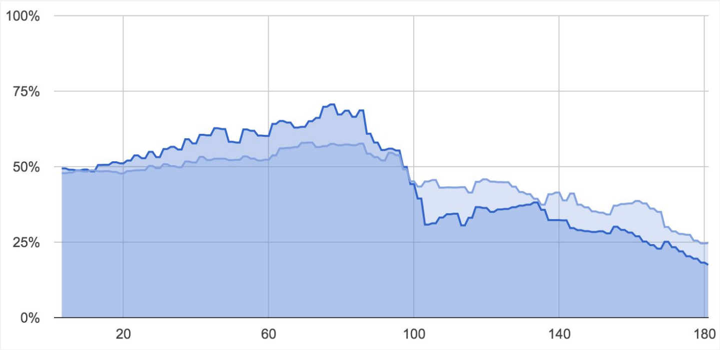
AlphaGo vs Lee Sedol: Game 3



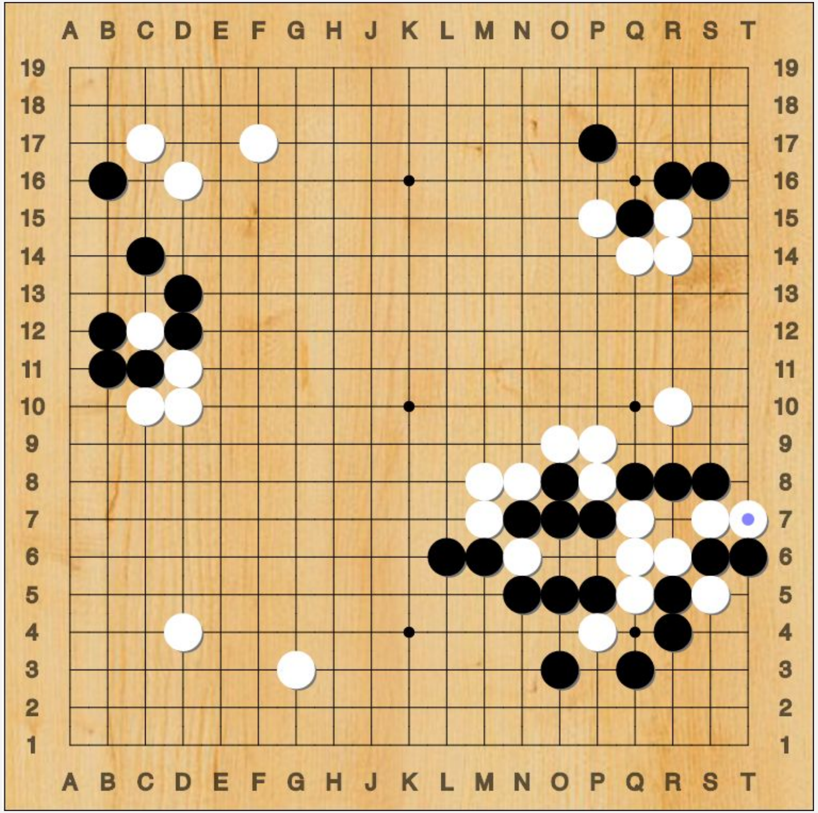


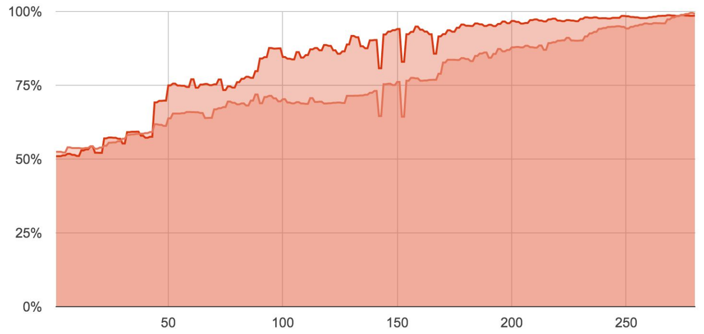
AlphaGo vs Lee Sedol: Game 4





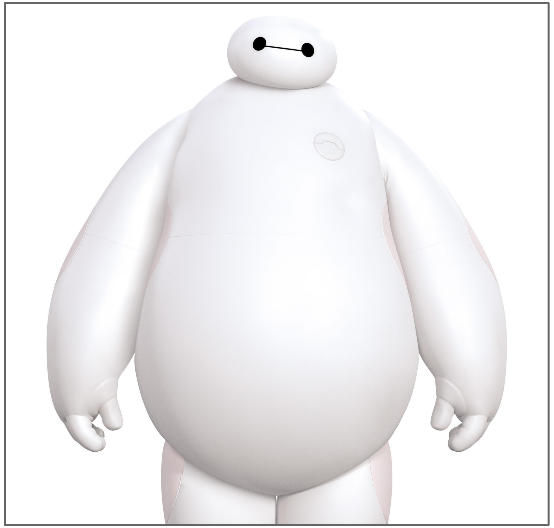
AlphaGo vs Lee Sedol: Game 5





Deep Blue	AlphaGo
Handcrafted chess knowledge	Knowledge learned from expert games and self-play
Alpha-beta search guided by heuristic evaluation function	Monte-Carlo search guided by policy and value networks
200 million positions / second	60,000 positions / second

What’s Next?



Google DeepMind

AlphaGo Team

With thanks to: Lucas Baker, David Szepesvari, Malcolm Reynolds, Ziyu Wang, Nando De Freitas, Mike Johnson, Ilya Sutskever, Jeff Dean, Mike Marty, Sanjay Ghemawat.

Google DeepMind