Mastering the game of Go without human knowledge

David Silver* Julian Schrittwieser* Karen Simonyan* · · ·

Google DeepMind -- Nature17

AntNLP -- Tao Ji

taoji.cs@gmail.com

Outline

- Go Description
- Motivation
- Neural network architecture
- Policy iteration
- Self-play reinforcement learning
- Monte Carlo tree search (MCTS)
- Neural network training
- Experiments

Description

Go

- `19–19=361`
- White `1`, Black `-1`, None `0`
- State: $s = (1, 0, -1, \ldots), \quad |s| = 361$
- Action: $a = (0, \ldots, 0, 1, 0, \ldots), \quad |a| = 361$

Motivation

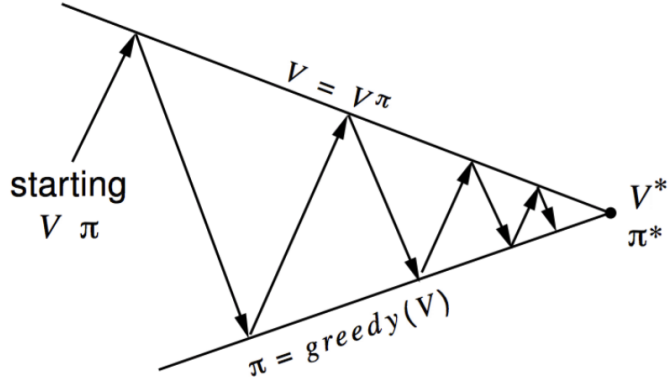
- It is trained solely by self-play reinforcement.
- Only use the black and white stones as input features.
- It uses a single neural network.
- It uses a simpler tree search.

Neural Network Architecture

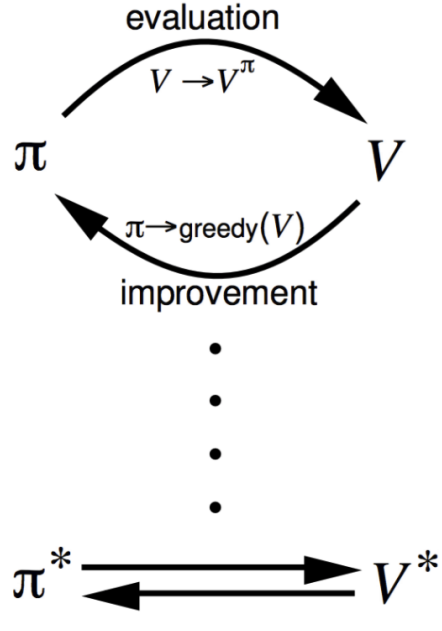
- Input: `19*19*17` (0/1 value)
- Output: $(\mathbf{p}, v) = f_\theta(s)$
  [move probabilities $\mathbf{p}$]: $p_a = \Pr(a|s)$
  [scalar evaluation v]: estimating the winning probability of s.
- Architecture: Residual Network with 20/40 Residual Block.

Generalized Policy Iteration

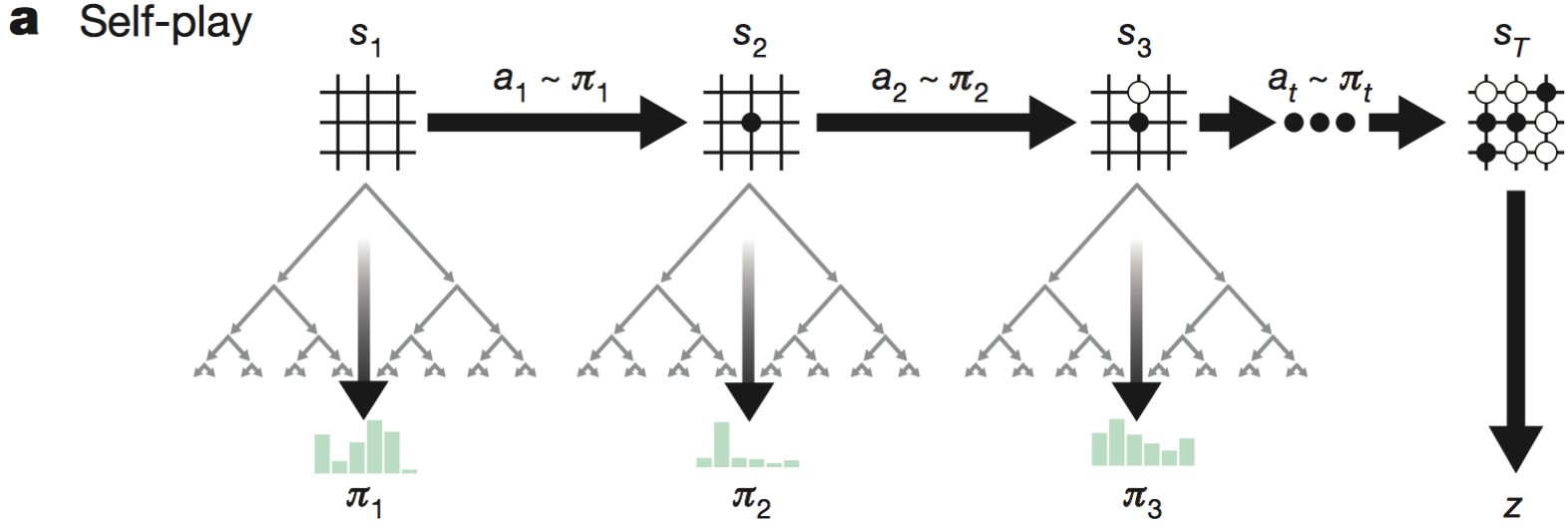
- Policy start from the π_0.
- Policy Evaluation: Get the value v_{π_0} of the π_0.
- Policy Improvement: Get policy π_1 based on v_{π_0}.
- Iterate steps 2 and 3 until get π^* and v^*.



Policy evaluation Estimate v_π
Any policy evaluation algorithm
Policy improvement Generate $\pi' \geq \pi$
Any policy improvement algorithm



Self-play Reinforcement Learning



Monte Carlo Tree Search (MCTS)

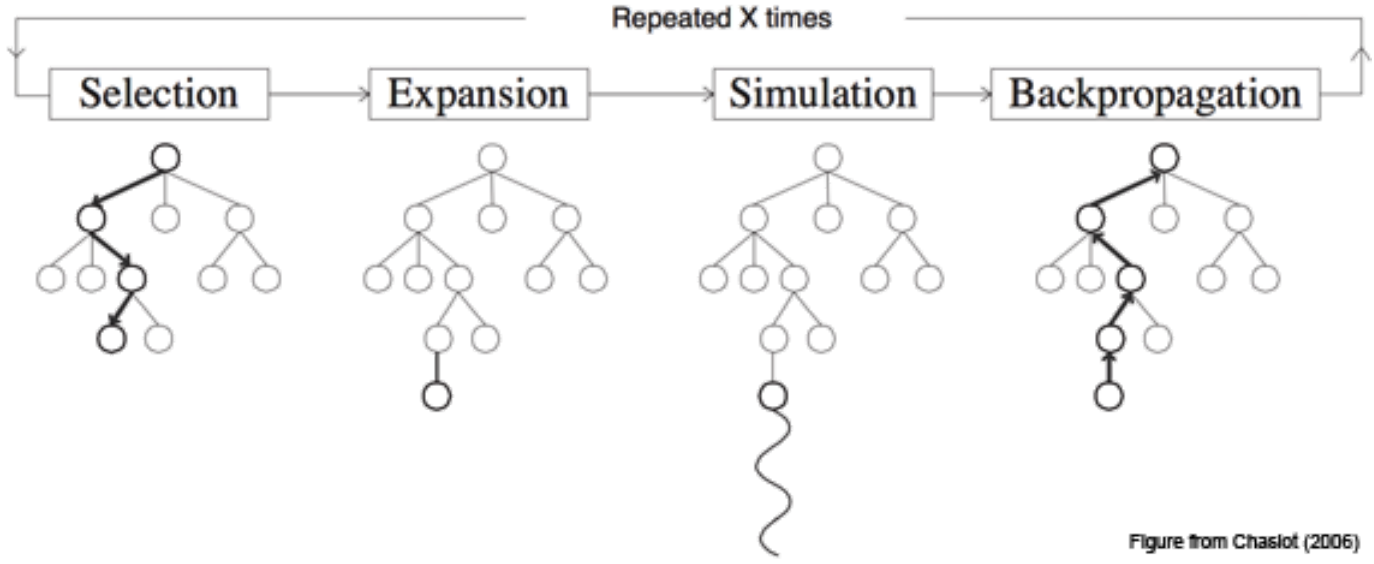
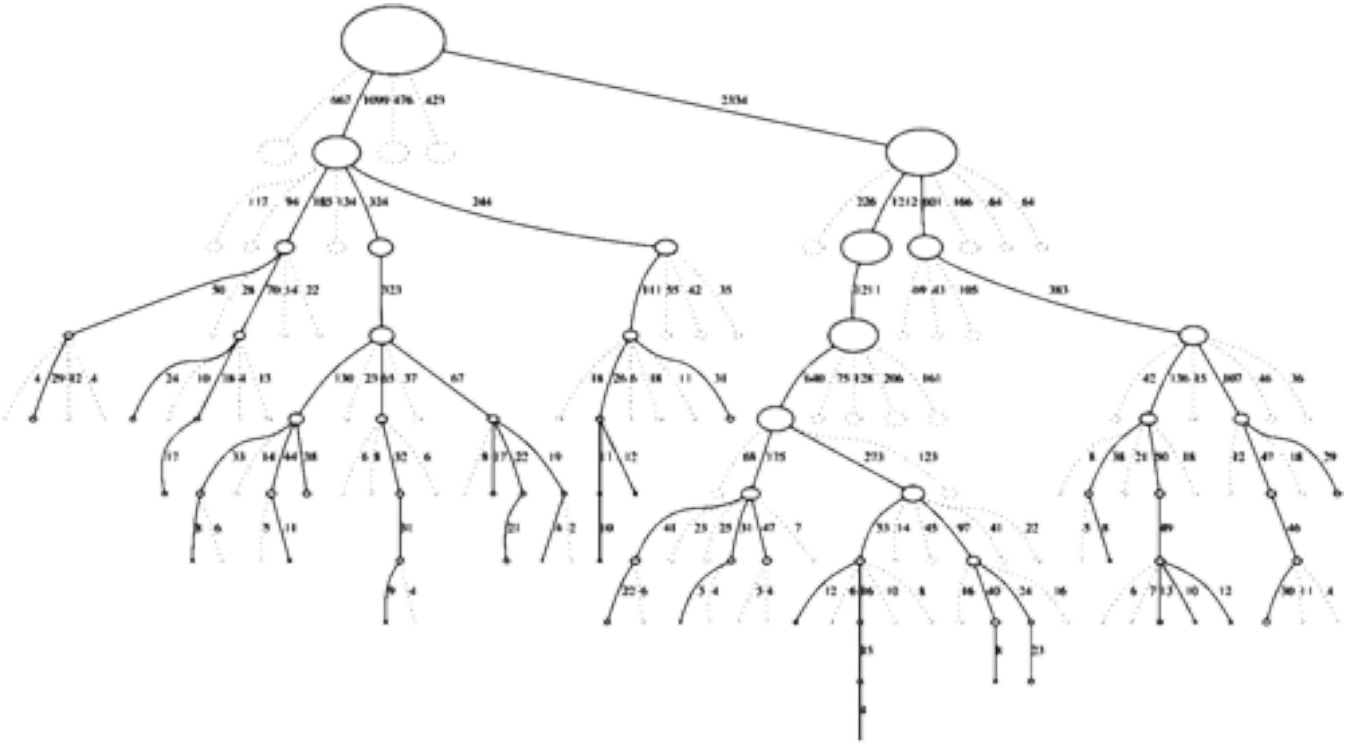
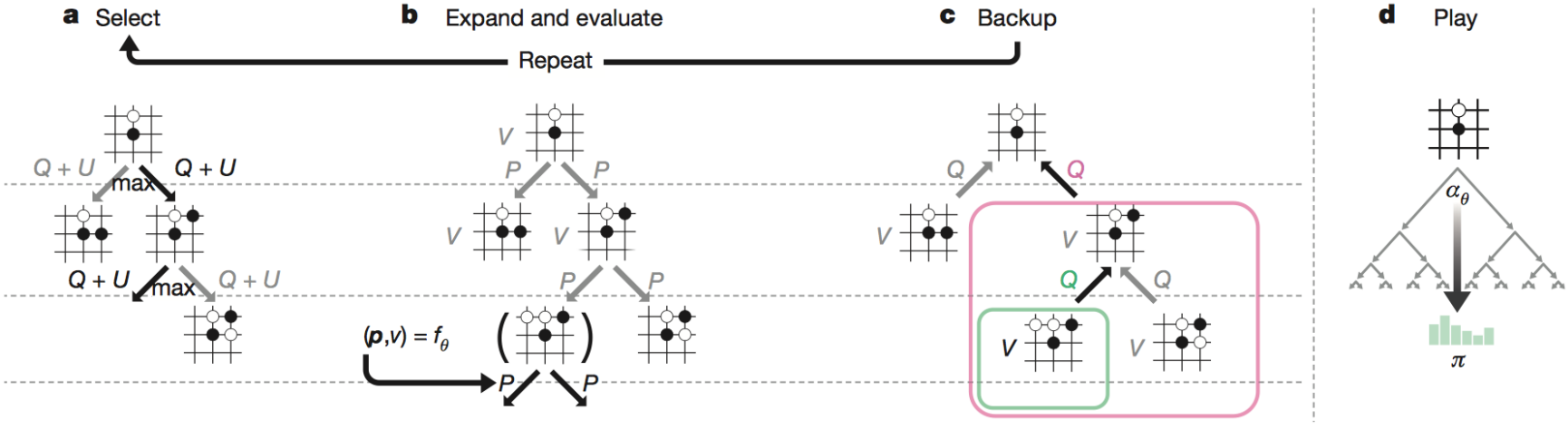


Figure from Chaslot (2006)

Monte Carlo Tree Search (MCTS)

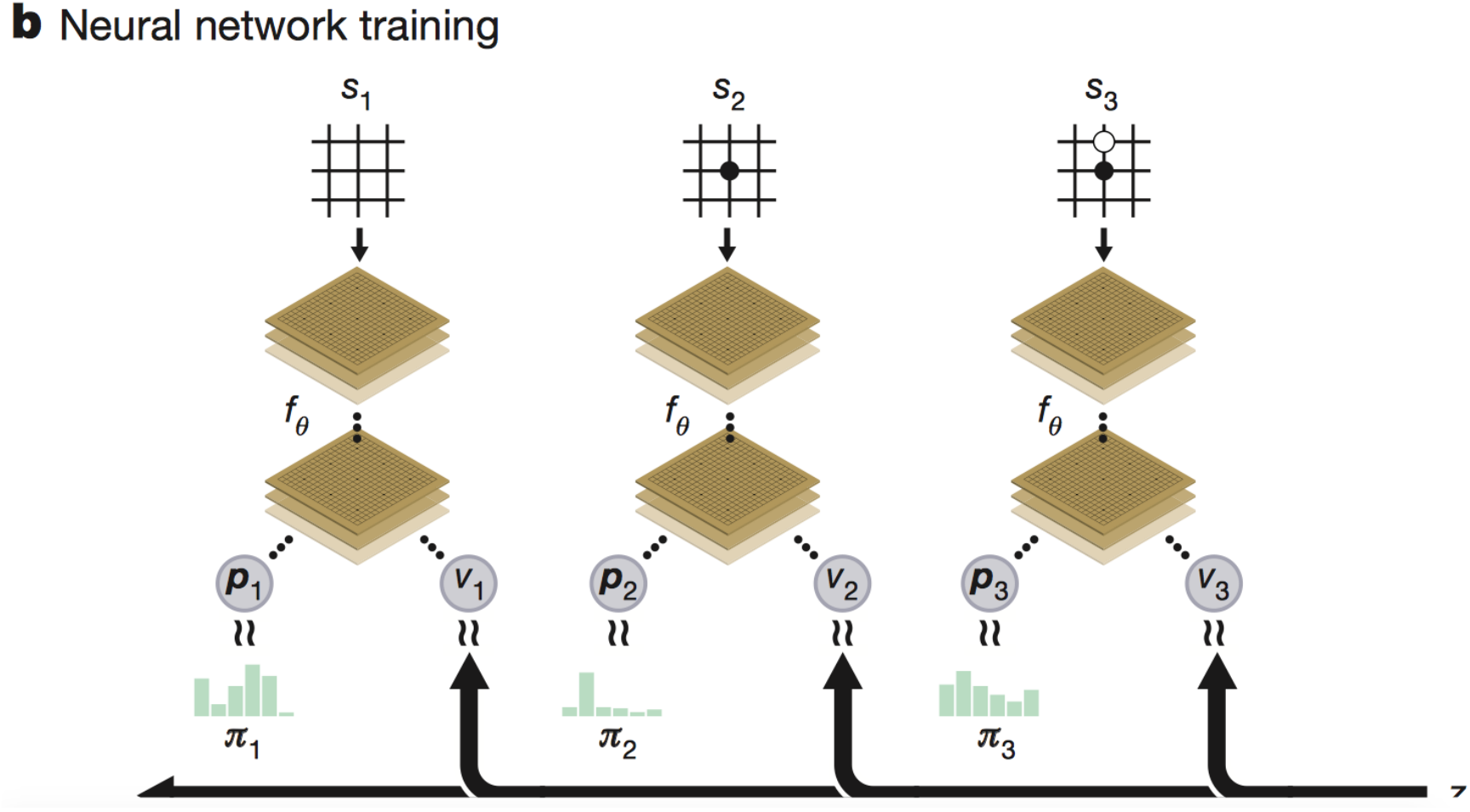


Monte Carlo Tree Search (MCTS)



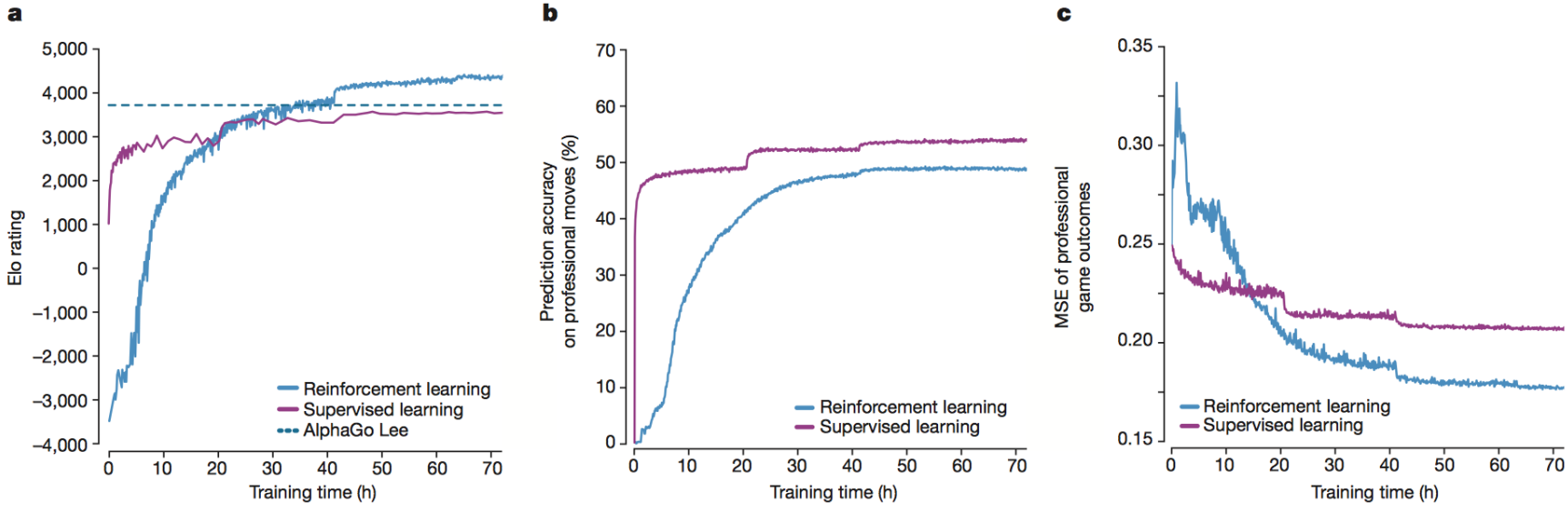
- $N(s,a), P(s,a)$
- $U(s,a) \propto \frac{P(s,a)}{1+N(s,a)}$
- $f_\theta(s') = (P(s',\cdot), V(s'))$
- $Q(s,a) = \frac{1}{N(s,a)} \Sigma_{s'|s,a \Rightarrow s'} V(s')$
- Upper confidence bound: $Q(s,a) + U(s,a)$
- $\pi(a|s) = \frac{N(s,a)^{1/\tau}}{\Sigma_b N(s,b)^{1/\tau}}$

Neural Network Training

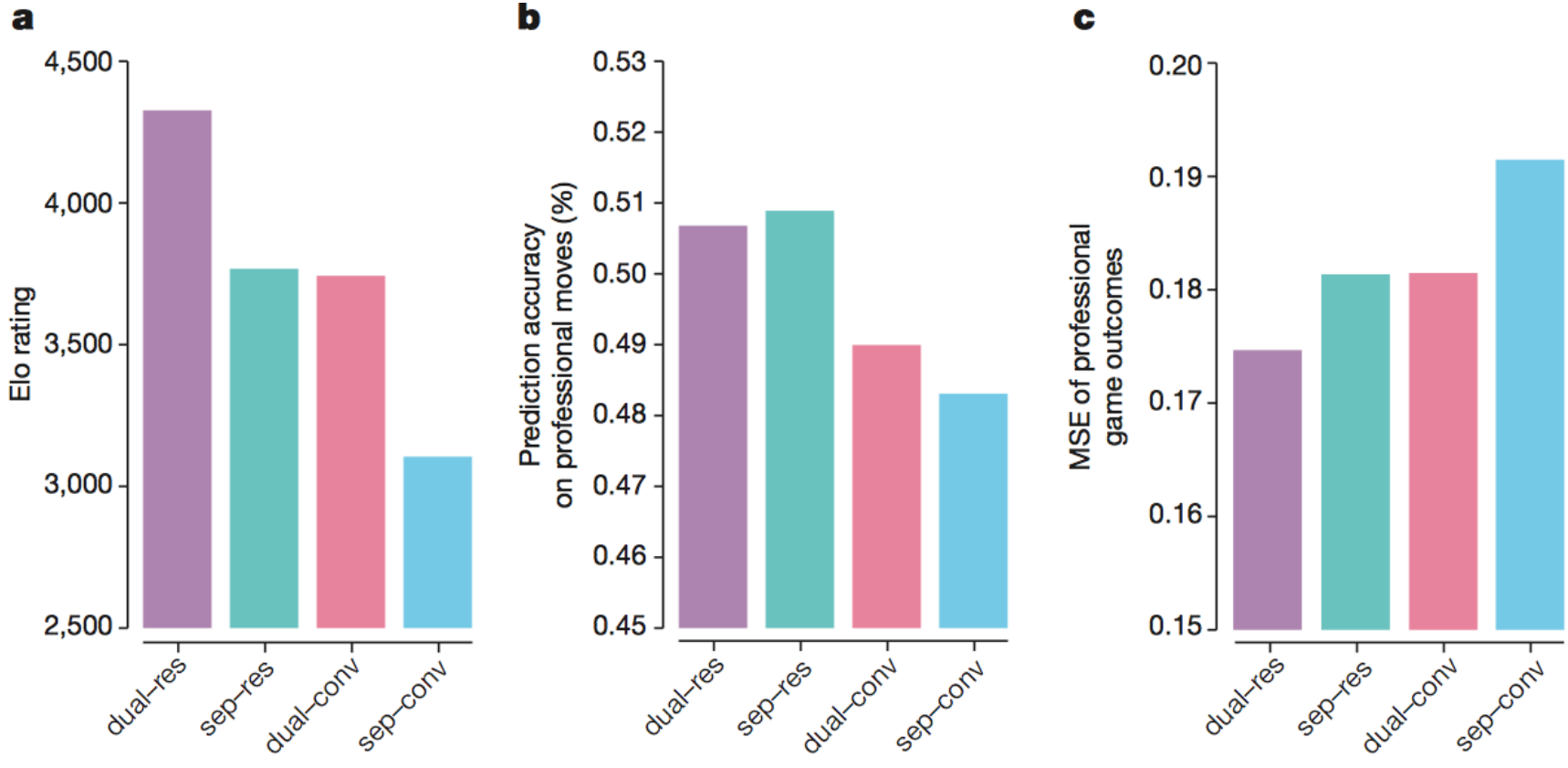


$$l = (z - v)^2 - \pi^T \log(\mathbf{p}) + c\|\theta\|^2$$

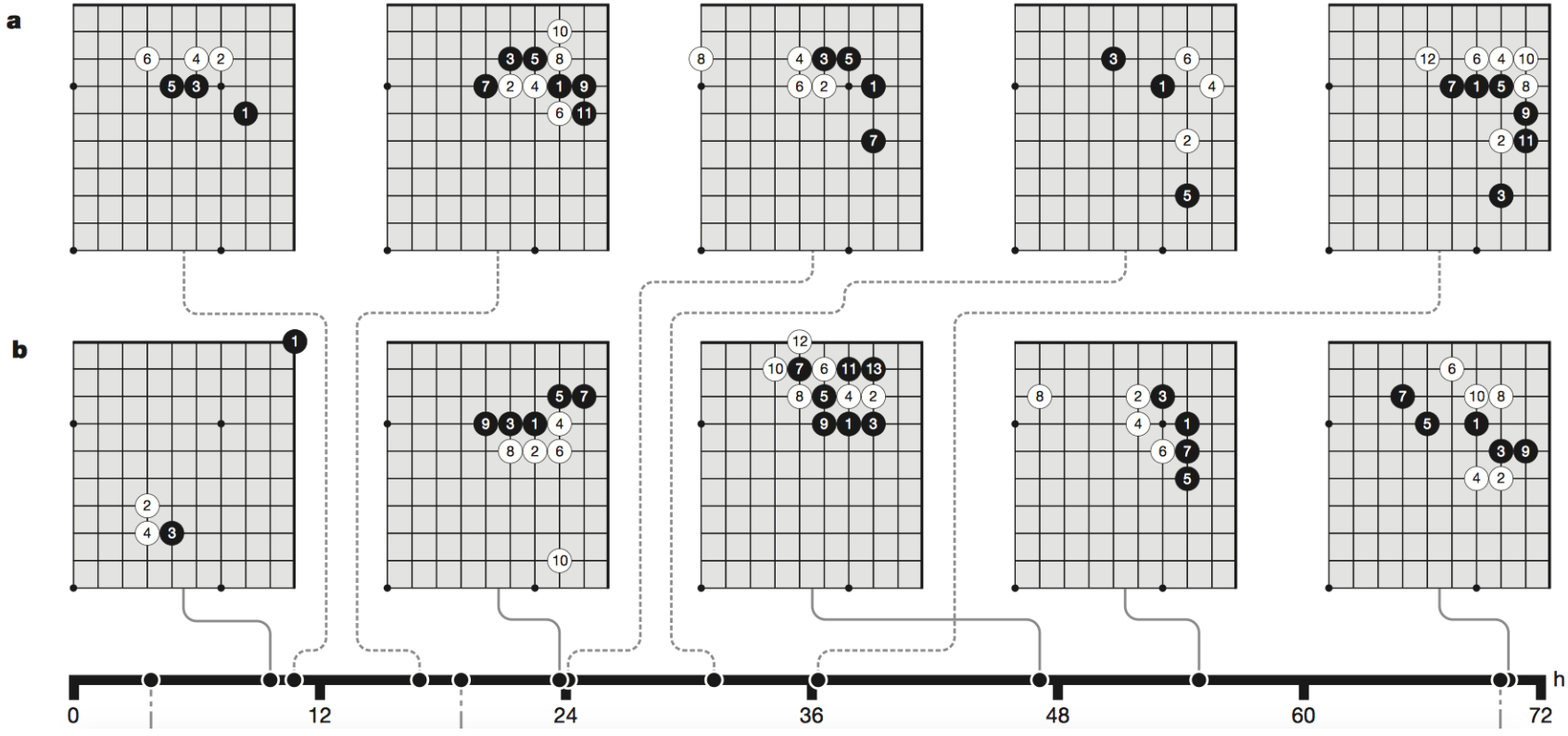
Experiments



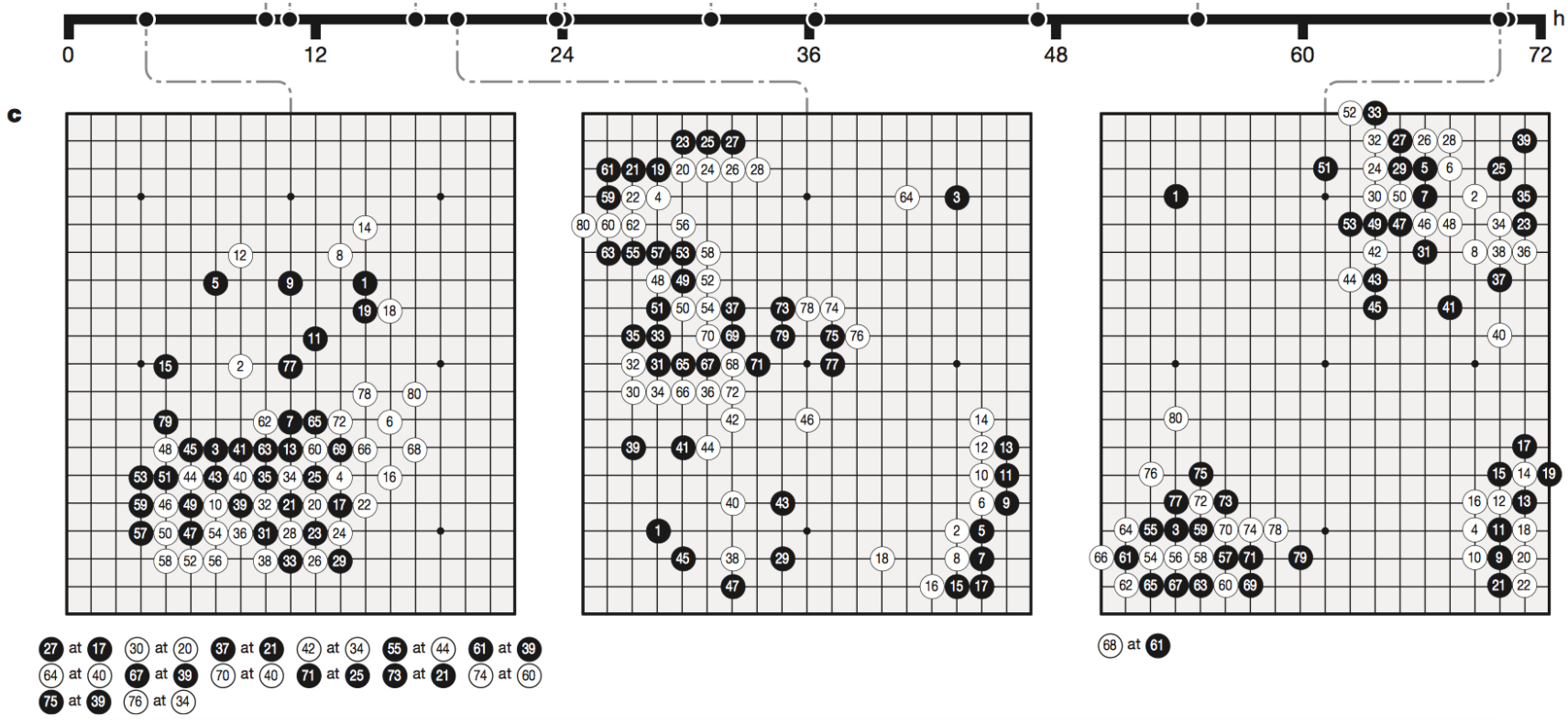
Experiments



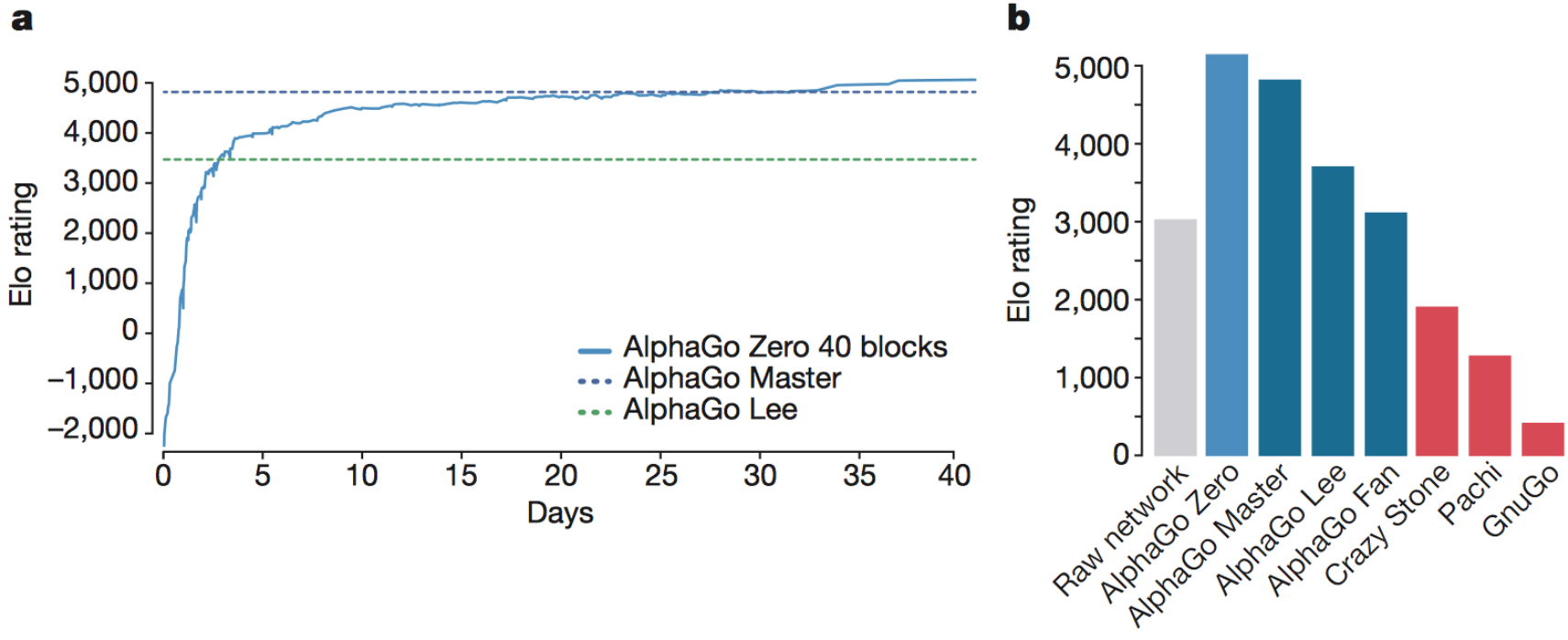
Experiments



Experiments



Experiments



Thank you!

Q&A