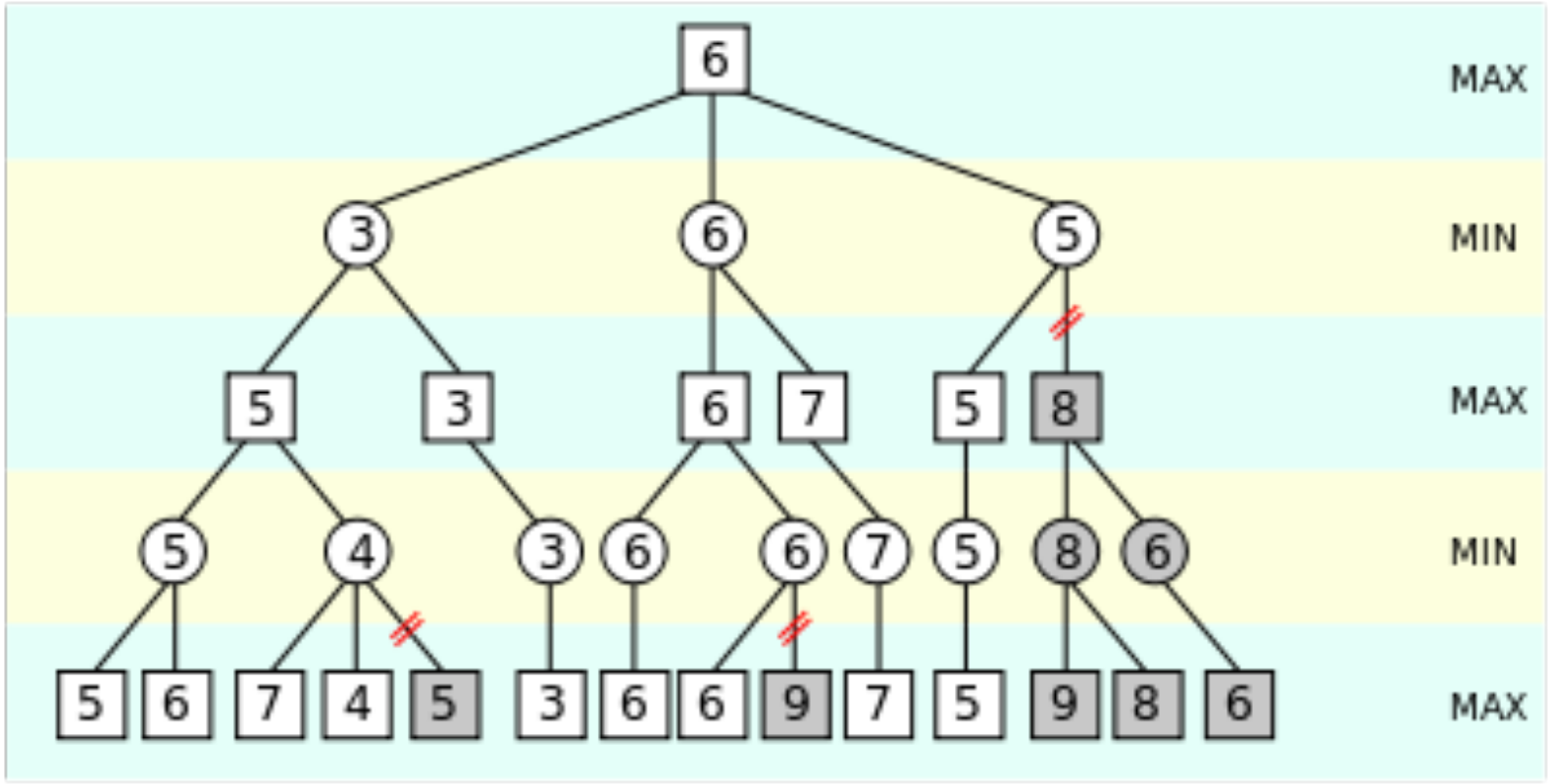
Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm

Alanova Shirin 193 group

Mechanical Turk

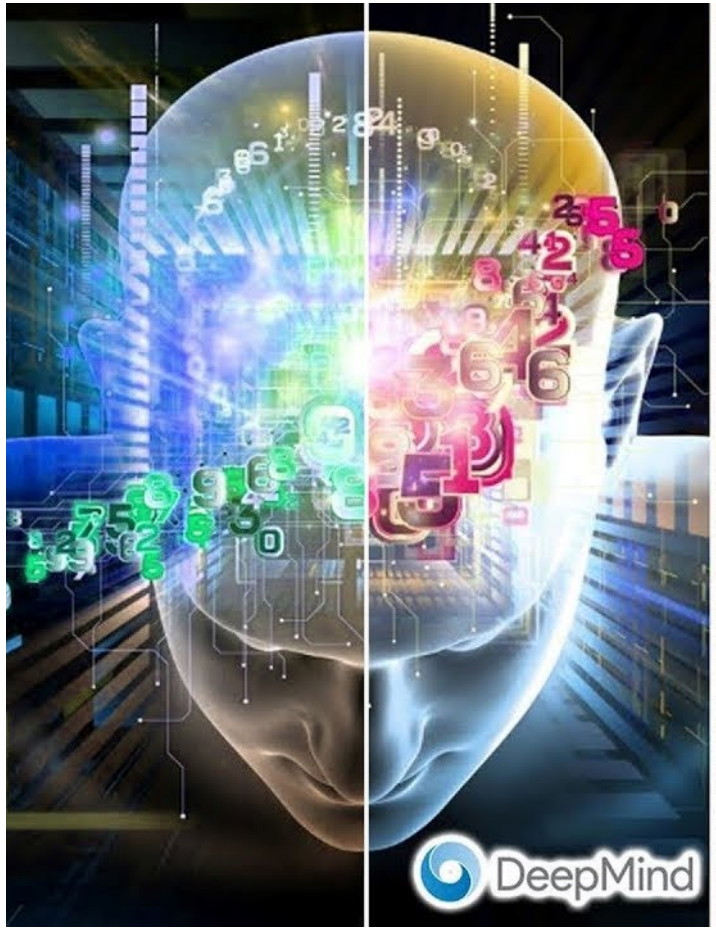
https://en.wikipedia.org/wiki/Mechanical_Turk

Alpha-beta search engine

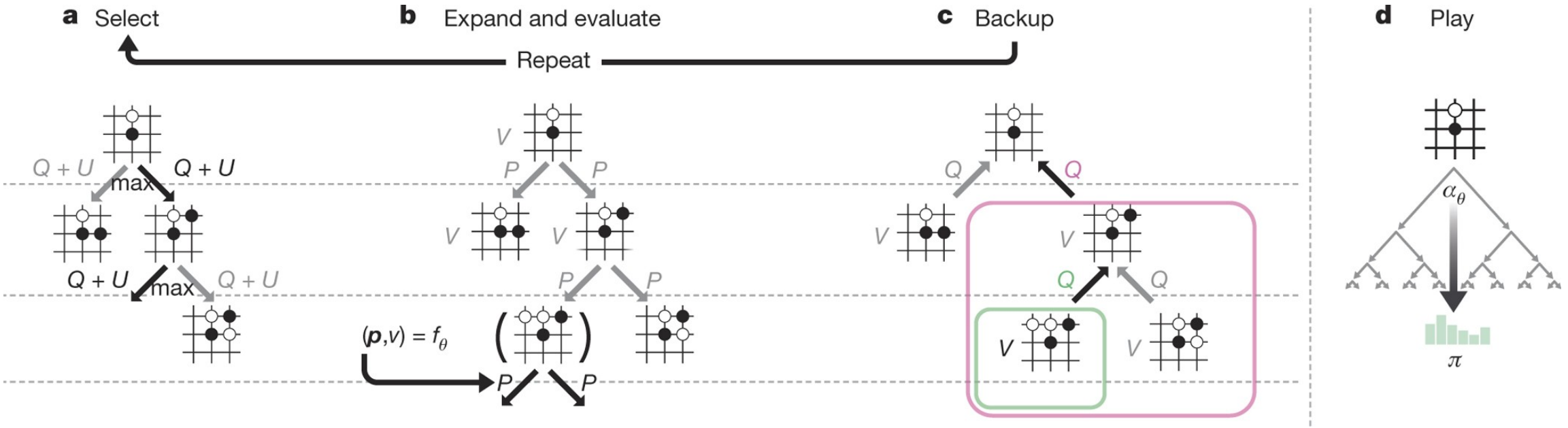


Deep Blue vs Kasparov (1997)

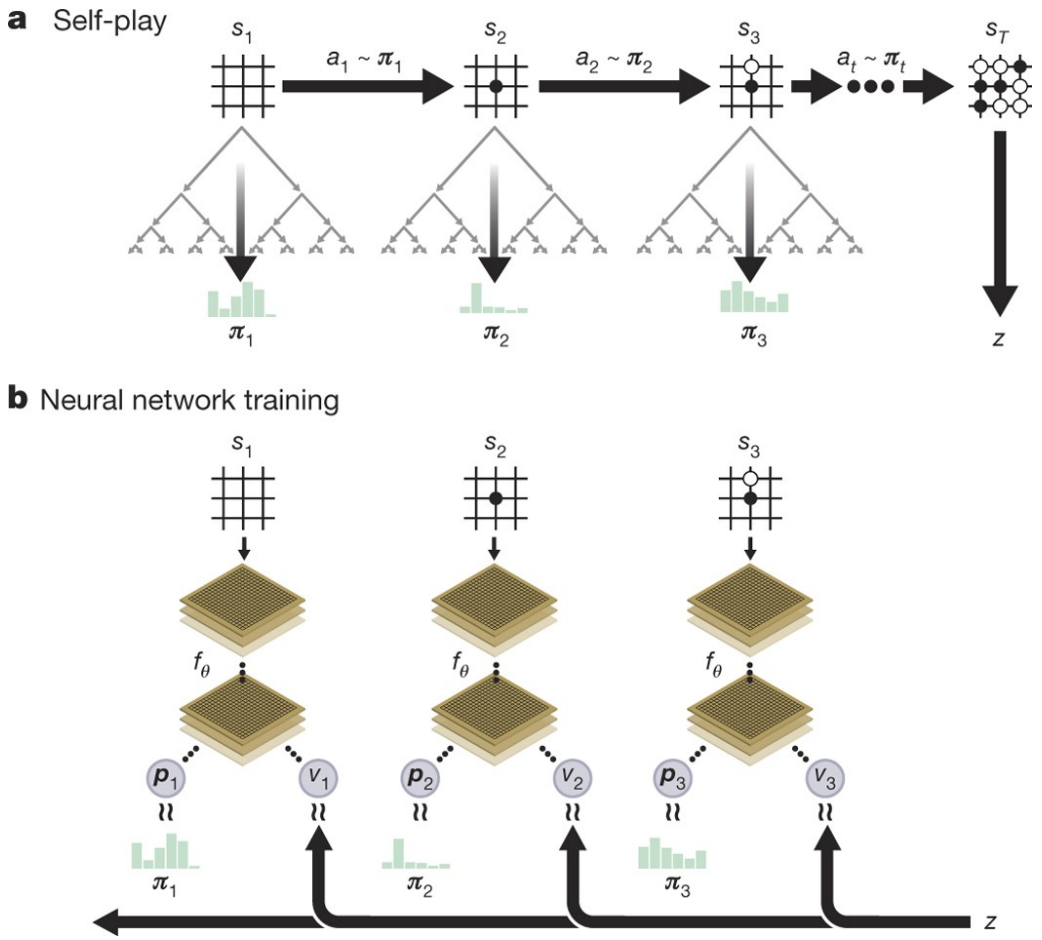
AlphaZero



MCTS in AlphaZero



AlphaZero architecture

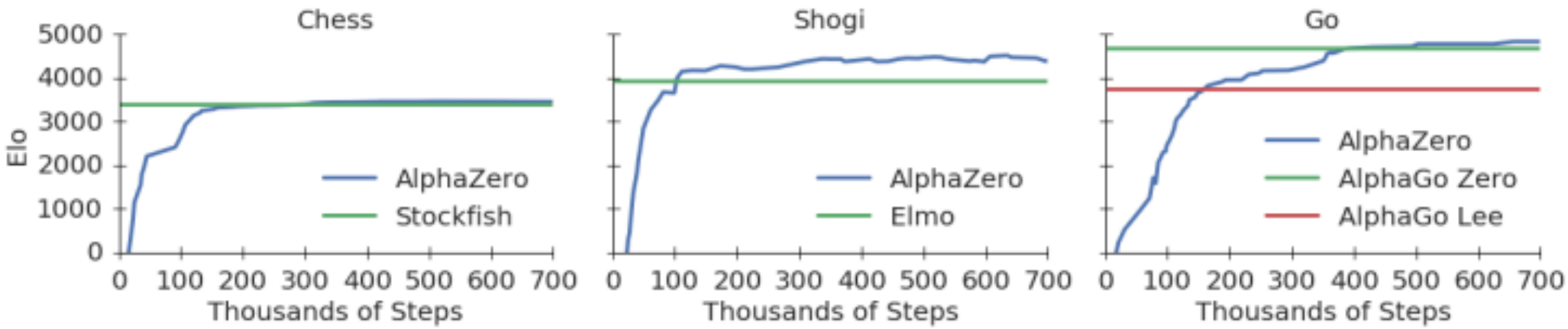


Loss function

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

AlphaGo Zero vs AlphaZero

AlphaZero during self-play reinforcement learning



Tournament evaluation of AlphaZero i

Game	White	Black	Win	Draw	Loss
Chess	*AlphaZero*	*Stockfish*	25	25	0
	Stockfish	*AlphaZero*	3	47	0
Shogi	*AlphaZero*	*Elmo*	43	2	5
	Elmo	*AlphaZero*	47	0	3
Go	*AlphaZero*	*AG0 3-day*	31	–	19
	AG0 3-day	*AlphaZero*	29	–	21

Table 1: Tournament evaluation of *AlphaZero* in chess, shogi, and Go, as games won, drawn or lost from *AlphaZero*'s perspective, in 100 game matches against *Stockfish*, *Elmo*, and the previously published *AlphaGo Zero* after 3 days of training. Each program was given 1 minute of thinking time per move.

Scalability of AlphaZero with thinking time, measured on an Elo scale

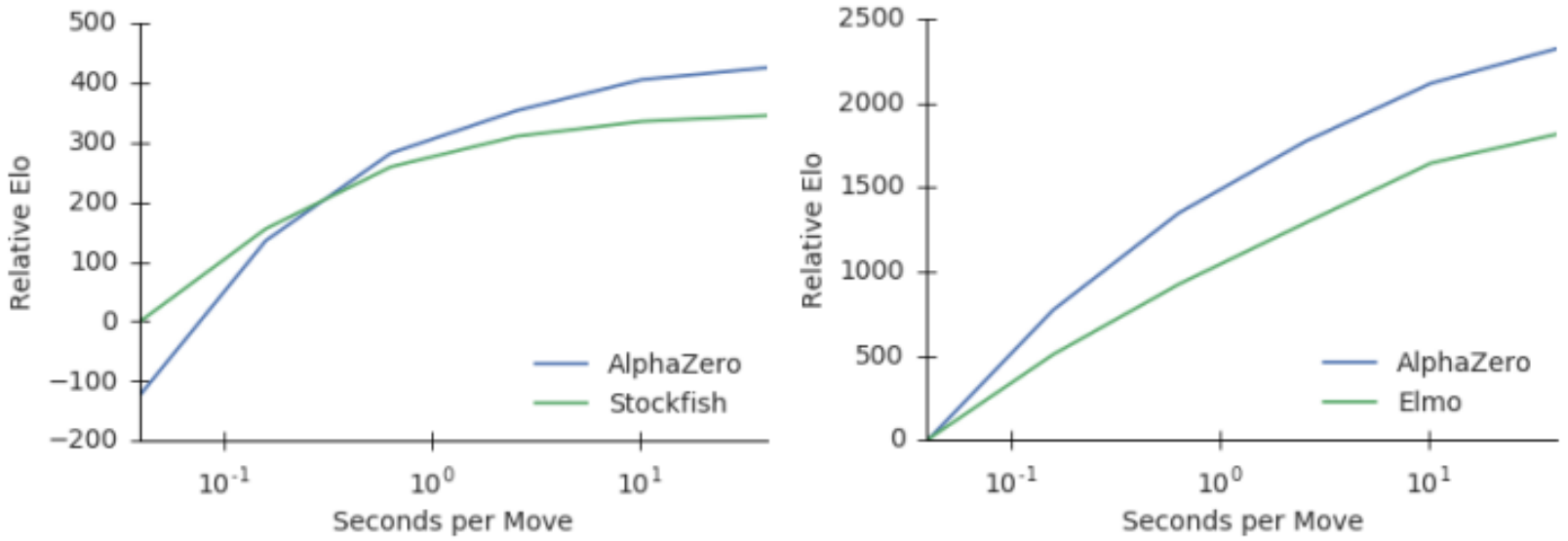
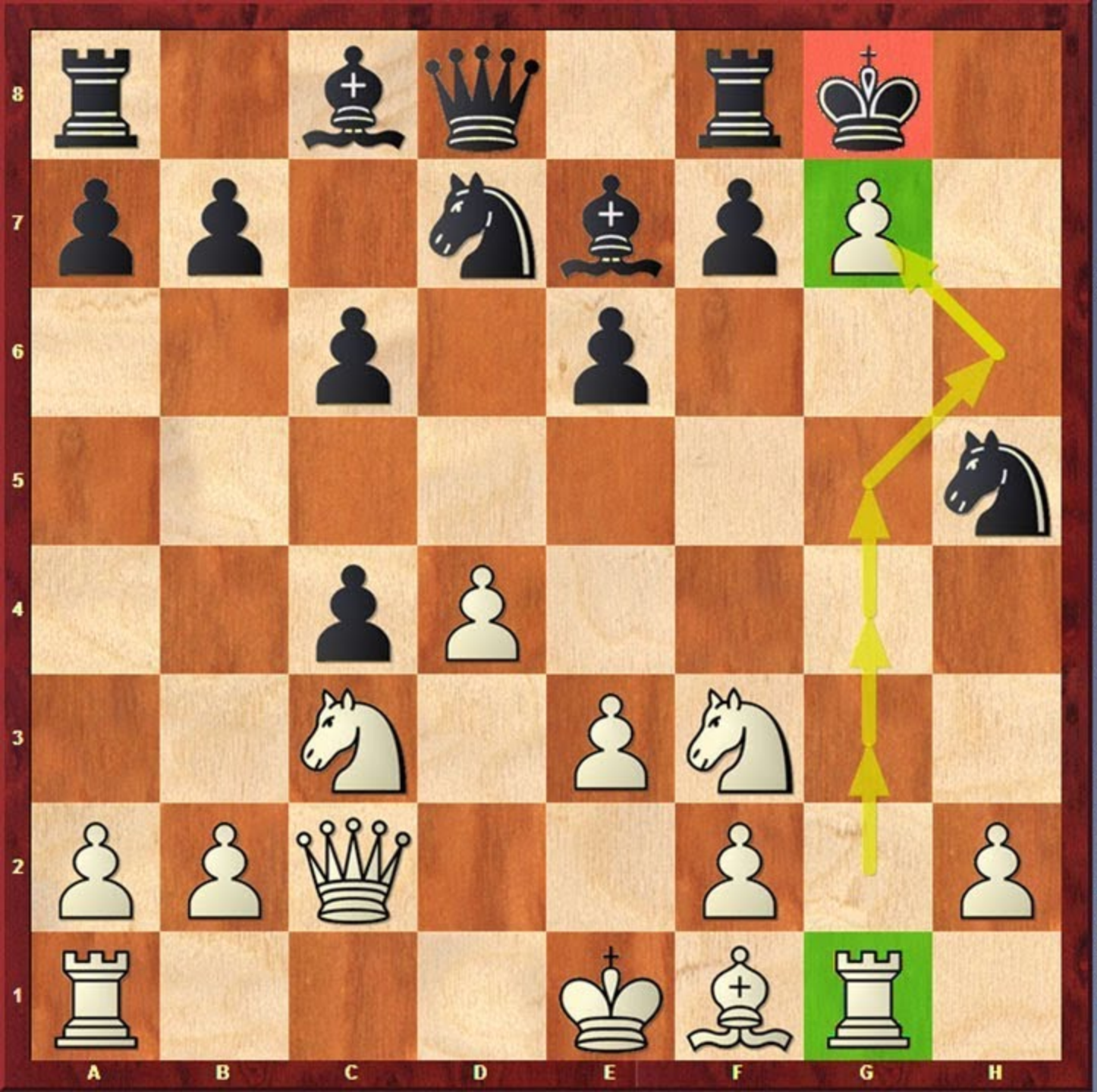
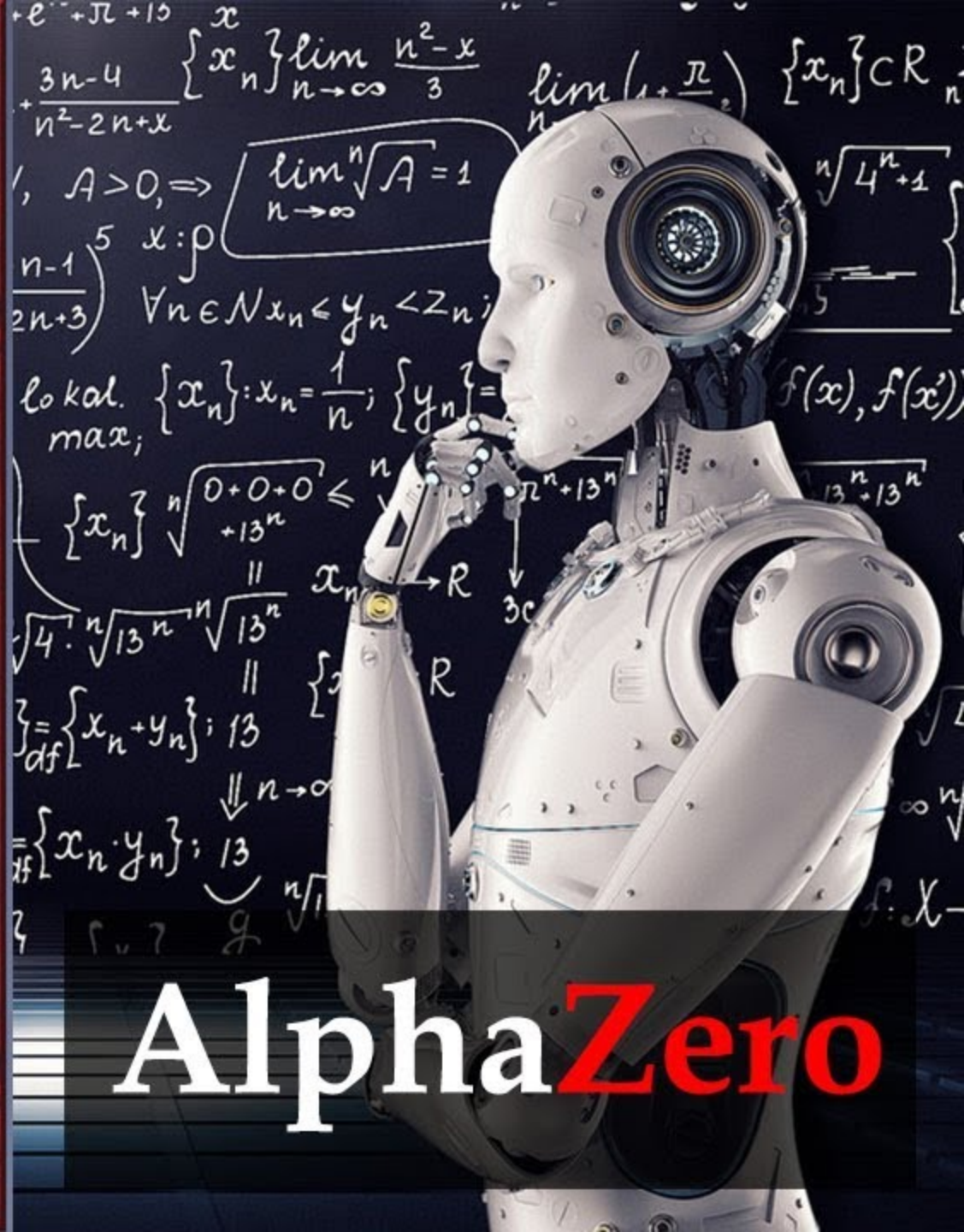


Figure 2: Scalability of *AlphaZero* with thinking time, measured on an Elo scale. **a** Performance of *AlphaZero* and *Stockfish* in chess, plotted against thinking time per move. **b** Performance of *AlphaZero* and *Elmo* in shogi, plotted against thinking time per move.


8
7
6
5
4
3
2
1
A
B
C
D
E
F
G
H


AlphaZero