

Playing Backgammon (Tesauro)

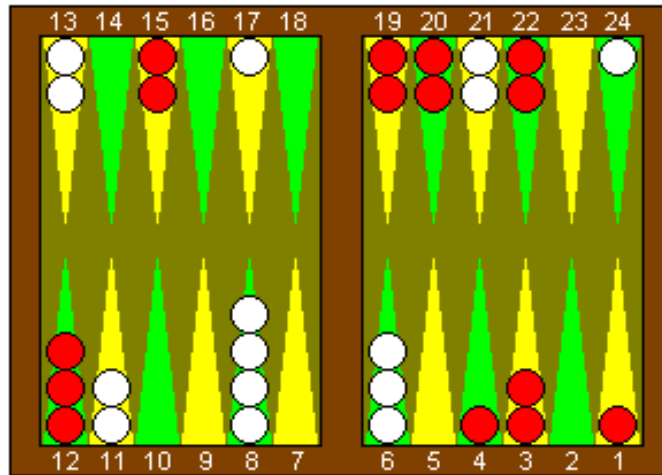
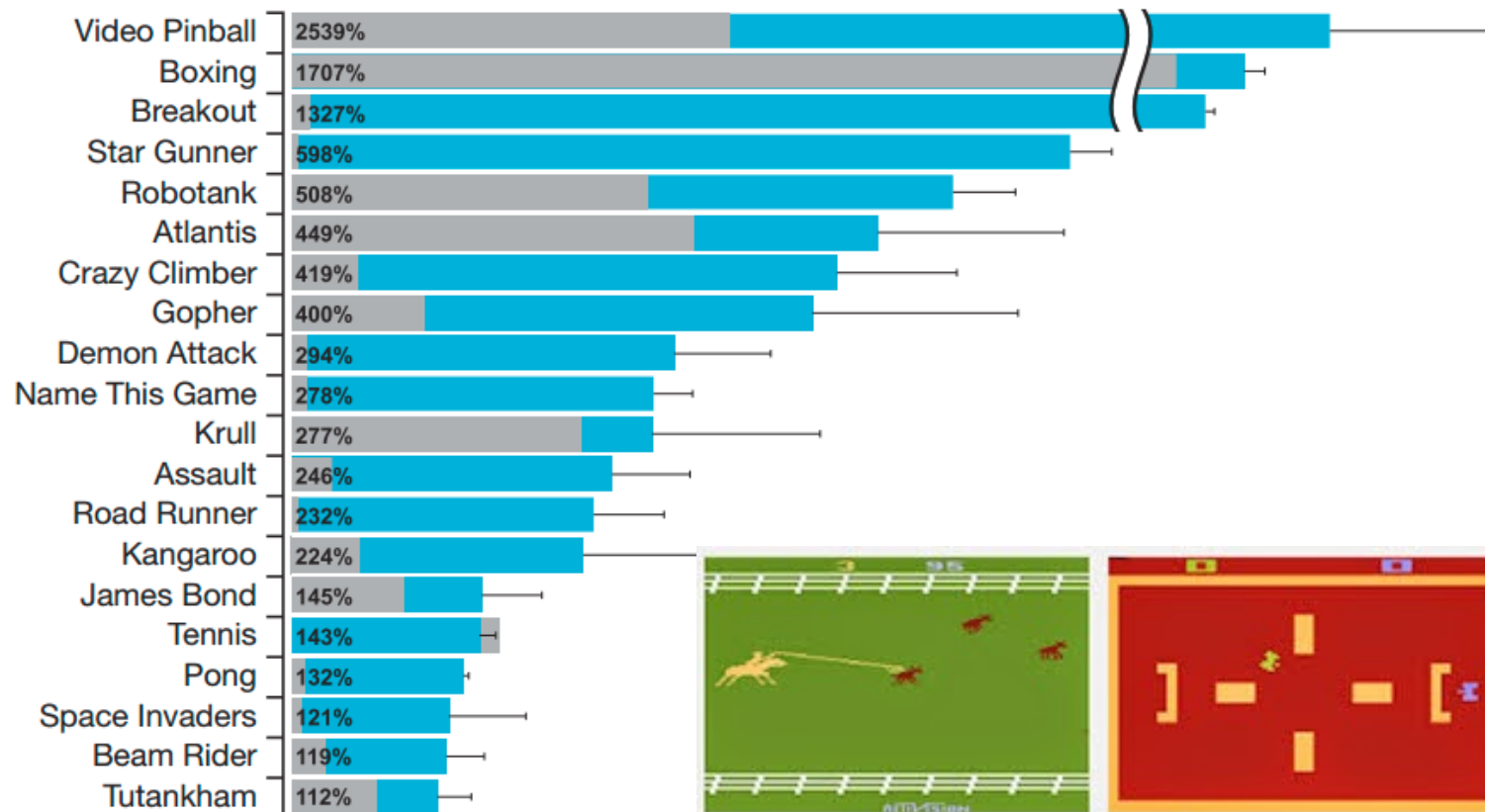


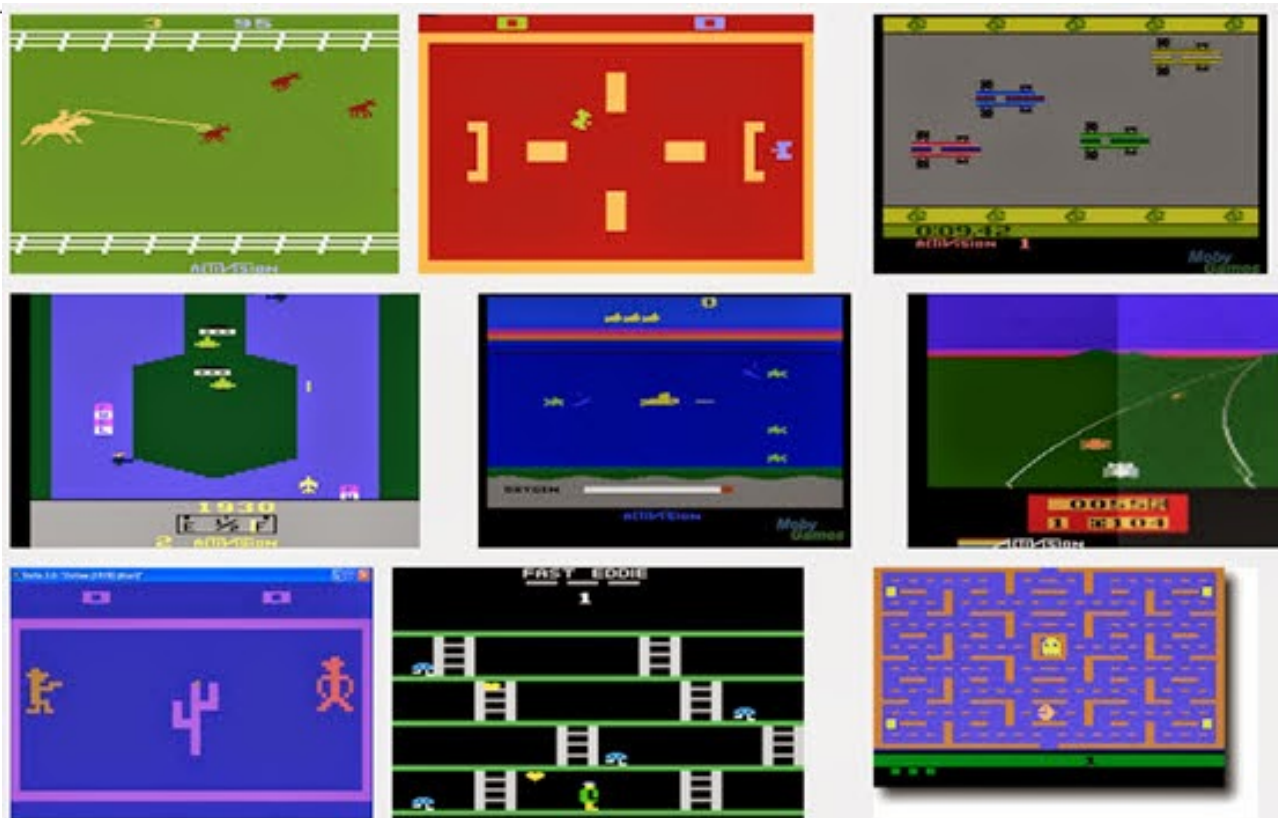
Figure 3. A complex situation where TD-Gammon's positional judgment is apparently superior to traditional expert thinking. White is to play 4-4. The obvious human play is 8-4*, 8-4, 11-7, 11-7. (The asterisk denotes that an opponent checker has been hit.) However, TD-Gammon's choice is the surprising 8-4*, 8-4, 21-17, 21-17! TD-Gammon's analysis of the two plays is given in Table 3.

Move	Estimate	Rollout
8-4*, 8-4, 11-7, 11-7	+0.184	+0.139
8-4*, 8-4, 21-17, 21-17	+0.238	+0.221

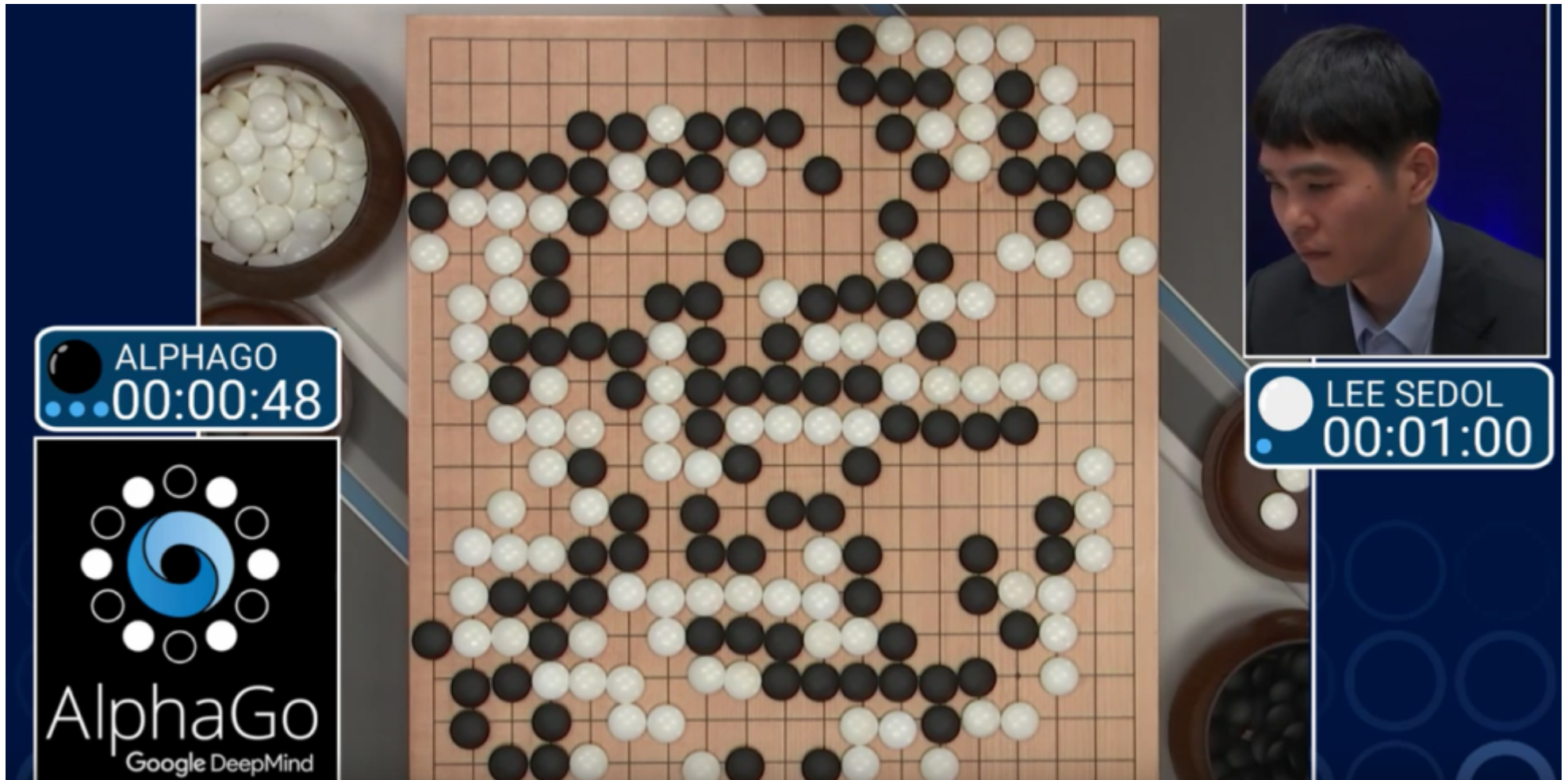
Table 3. TD-Gammon's analysis of the two choices in Figure 3. The estimated equity is the neural network's output at the 1-ply level (i.e., no lookahead). The rollout is actual outcome playing each position out 10,000 times to completion with different random dice sequences (see the appendix). Standard deviation in the rollout results is approximately 0.01.



Playing Atari (DeepMind)



Playing Go



<http://pop.h-cdn.co/assets/16/10/980x490/landscape-1457623728-alphago2.png>