



LEE JIN-MAN/AP/REX/SHUTTERSTOCK

Figure 1 | AlphaGo versus Lee Sedol. In March 2016, the artificial-intelligence program AlphaGo defeated a world Go champion, Lee Sedol.

FORUM Artificial intelligence

Learning to play Go from scratch

An artificial-intelligence program called AlphaGo Zero has mastered the game of Go without any human data or guidance. A computer scientist and two members of the American Go Association discuss the implications. [SEE ARTICLE P.354](#)

A big step for AI

SATINDER SINGH

When chess fell to computers¹, Go was left standing as the board game that humans could count on to dominate computers for a long time. In a result that surprised many at how soon it arrived, the artificial-intelligence (AI) program AlphaGo² defeated a world Go champion, Lee Sedol, in 2016 (Fig. 1). AlphaGo built on earlier work^{3–5} and was a fantastic accomplishment for AI, but there was one important caveat: its training required the use of expert human gameplay. On page 354, Silver *et al.*⁶ report an updated version of the program, AlphaGo Zero, that uses a method called reinforcement learning, free of human guidance. The AI massively outperforms the already superhuman

AlphaGo and, in my view, is one of the biggest advances, in terms of applications, for the field of reinforcement learning so far.

How does AlphaGo Zero work? It uses the current state of the game board as the input for an artificial neural network. The network calculates the probability with which each possible next move could be played and estimates the probability of winning for the player whose turn it is to make the move. The AI learns the moves that will maximize its chance of winning through trial and error (reinforcement learning) and was trained exclusively by playing games against itself.

During training, AlphaGo Zero used about 0.4 seconds of thinking time per move to perform a look-ahead search — that is, it used a combination of game simulations and the outputs of its neural network to decide which moves would give it the highest probability of winning. It then used this information

to update its neural network. Although the above is a simplified description of Silver and colleagues' reinforcement-learning method, it highlights how intuitive and straightforward it is compared with the approach used by AlphaGo, which required many neural networks and multiple sources of training data.

How well did AlphaGo Zero do? There was roughly an order of magnitude improvement in most of the relevant numbers for AlphaGo Zero compared with those for the version of AlphaGo² that defeated Lee Sedol: 4.9 million training games versus 30 million training games, 3 days of training versus several months of training, and a single machine that has 4 tensor processing units (TPUs; specialized chips for neural-network training) versus multiple machines and 48 TPUs. Playing under conditions that match those of human games, AlphaGo Zero beat AlphaGo 100–0.

So, what does this all mean? First, let's consider this question in terms of the field of reinforcement learning. The improvement in training time and computational complexity of AlphaGo Zero relative to AlphaGo, achieved in about a year, is a major achievement. Although the authors' training method is new, it combines some basic and familiar aspects of reinforcement learning. Taken together, the results suggest that AIs based on reinforcement learning can perform much better than those that rely on human expertise. Indeed, AlphaGo Zero will probably be used by human Go players to improve their gameplay and to gain insight into the game itself.

Second, let's consider what the results mean for the media obsession with AI versus humans. Yes, another popular and beautiful game has fallen to computers, and yes, the authors' reinforcement-learning method will be applicable to other tasks. However, this is not the beginning of any end because AlphaGo Zero, like all other successful AI so far, is extremely limited in what it knows and in what it can do compared with humans and even other animals.

Satinder Singh is in the Computer Science and Engineering Department, University of Michigan, Ann Arbor, Michigan 48109, USA. e-mail: baveja@umich.edu

Conversations with AlphaGo

ANDY OKUN & ANDREW JACKSON

Edward Lasker, a chess grandmaster and Go enthusiast, is reported to have said that "the rules of Go are so elegant, organic and rigorously logical that if intelligent life forms exist elsewhere in the Universe, they almost certainly play Go". In some sense, Silver and colleagues' work proves Lasker's hypothesis — it demonstrates that an inhuman intelligence plays Go in a way that is somewhat similar to human players.

The rules of Go could hardly be simpler, yet the complexity that emerges is dizzying. Human players grapple with this complexity partly by analysis: studying tactics, memorizing established patterns and learning to probe deeply into the coming moves. Professional players, who compete for millions of dollars in prize money, train from as early as four years old to master these skills. Their attainment is extraordinary — thinking a hundred moves ahead and accurately assessing the board at a glance is *de rigueur*. But analysis is just the foundation. Go players also have to accrue a body of wisdom and experience, rules of thumb, proverbs, strategic concepts and even a feel for the shapes that the stones (playing pieces) make.

Put simply, they require judgement and intuition to play well.

AI has now met, and exceeded, the skill of the best human players. In doing so, it has posed the question of how much we really know about the game. A legendary Go player — one who changes our conceptions of the game — might come along only once in a century. When AlphaGo defeated Lee Sedol 9p (9p is the top level of accomplishment in Go), were we meeting the next legend? And would we have to throw away centuries of lore and study?

Earlier this year, an updated version of AlphaGo called AlphaGo Master played and won 60 games against top professionals. These games are still being dissected by players and fans everywhere. An additional 50 games that AlphaGo Master played against itself, released after the AI defeated the current world number one, Ke Jie 9p, are also being mined for insights into the AI's choices, particularly its opening moves.

AlphaGo Zero will now provide the next rich vein. Its games versus AlphaGo Master will surely contain gems, especially because its victories seem effortless. At each stage of the game, it seems to gain a bit here and lose a bit there, but somehow ends up slightly ahead, as if by magic. The AI's self-play games, like those of AlphaGo Master, are all-out brawls, as one would expect from two players whose judgements are identical — in perfect agreement on the stakes, neither player can give an inch.

Silver and colleagues' results suggest that centuries of human gameplay have not been wholly wrong. AlphaGo Zero independently found, used and occasionally transcended

many established sequences of moves used by human players. In particular, the AI's opening choices and end-game methods have converged on ours — seeing it arrive at our sequences from first principles suggests that we haven't been on entirely the wrong track. On the other hand, some of its middle-game judgements are truly mysterious and give observing human players the feeling that they are seeing a strong human play, rather than watching a computer calculate.

Go players, coming from so many nations, speak to each other with their moves, even when they do not share an ordinary language. They share ideas, intuitions and ultimately, their values, over the board — not only particular openings or tactics, but whether they prefer chaos or order, risk or certainty, and complexity or simplicity. The time when humans can have a meaningful conversation with an AI has always seemed far off and the stuff of science fiction. But for Go players, the day is here. ■

Andy Okun and Andrew Jackson are in the American Go Association, New York, New York 10163-4668, USA.

e-mails: president@usgo.org; andrew.jackson@usgo.org

1. Campbell, M., Hoane Jr, A. J., Hsu, F.-H. *Artif. Intell.* **134**, 57–83 (2002).
2. Silver, D. *et al. Nature* **529**, 484–489 (2016).
3. Tesauro, G. *Commun. ACM* **38** (3), 58–68 (1995).
4. Silver, D., Sutton, R. S. & Müller, M. *Mach. Learn.* **87**, 183–219 (2012).
5. Gelly, S. *et al. Commun. ACM* **55** (3), 106–113 (2012).
6. Silver, D. *et al. Nature* **550**, 354–359 (2017).

A.J. declares competing financial interests: see go.nature.com/2yj8c5d for details.