Review of: "Self-Replication, Spontaneous Mutations, and Exponential Genetic Drift in Neural Cellular Automata"

Kieran Greer

Potential competing interests: No potential competing interests to declare.

I think that this paper shows very good ability to implement some code, but the test scenario itself may be erronous. I think that the testing has become confused, especially in the results section 'Genetic Encoding and Exponential Drift', for example. You state several changes from the standard procedure - that a mutation should move the state closer to the goal. Is it possible that these changes allow the state to move further away from the goal as well and therefore, if you start with something that is close to the goal, mutations can change it in the other direction? You note that ALife prefers divergence, but then it would not be a surprise if the evolution paths diverged.

You state a simplistic model with only 70 pixels or so, but then your evolutions do many different things. You only note the pixel colour as part of the coding, but then a fish, for example, can grow new stripes or is less able to produce an egg. How do you create a stripe from a pixel change? So this must be part of the DNA coding as well.

On page 6, this suggests a much more sensitive system:
'However, this solution is still unsatisfying, as it could well be closer to “paint by numbers” than to a modular genetic code. In the worst case, a fish’s appearance could in theory be fully decided by one out of hundreds of bit being 0 in one egg, and all other values could lead to a lizard. The current training method does not guarantee that the NCA will not use DNA as a discrete identifier.'

The section Calculating the Genetic Drift:
'We define genetic drift as the accumulation of neutral mutations in the genetic code through successive generations. ....'

What exactly is a neutral mutation. If there is no selection based on a fitness function, the number of mutations that would move the state closer to the goal is much less than all possible moves, so over time with random mutations, you would expect the number of neutral moves to be much larger.

Where or when is the NCA used, because it should tend to convergence. Again, you note this in the Discussion, where you describe a divergence change to it, so is there a consistent method through the whole paper?

From A to 2A with possibly A*. OK, a small difference is understood, but why should it go from A to 2A? The NCA is trained to duplicate A, but any errors in A are then exaggerated as well, so maybe you need the target or fitness function here.

There are some grammar mistakes and missing words, so please check this.