

## Homework 1

Cornel de Vroomen

What actually was done

Computers can sometimes be tricky to fully understand. Nowadays with new technologies, there are more and more “black boxes”, which are programs which we give input and gives output. What happens in between is not totally clear. At Deepmind in London, a computer program is created which is able to predict what other intelligent computer programs will do now, and will do in the future.

What was the experiment

The experiment was to predict what intelligent characters who are trying to get boxes in a virtual room would do. These characters are learning to try to find as much boxes as possible, because a box stands for points. There were three different characters who would all have a different initial strategy. After looking at the characters running around, the ToMnet program was able to predict what step the different characters were going to make in the future.

What ideas from ToM were used

Theory of Mind means that we as humans can predict what other humans are thinking, or what their mental state is. In this research outcome, it is all about predicting what the mental state of the computer program is, and to try to find out what they are thinking of. Before we only had to predict in what mental state intelligent creatures named ‘human’ are in, but now there are new intelligent creatures in town: the computer.

What is the solution architecture

The solution architecture is an interesting one. There were 3 neural networks used which work together, and which is being compared to the human brain. The first neural network consisted of trying to learn what the characters are doing based on their past performance. The second neural network tries to understand their current state. Lastly, the third neural network tried to think of the future behavior for the characters.