Three Expectations of Intelligence

by Sven Nilsen, 2017

The default three expectations of intelligence is the following:

1. Making equal or better decisions on average than a random decision.
2. Making equal or better decisions on average than doing nothing.
3. Making no worse decisions on average than self-predicted, when imagined given more time.

Making random decisions and doing nothing are fixed rules that fit in every kind of agent/environment interactions. This is why they are listed here and not any arbitrary fixed rule.

Sometimes, making a random decision is optimal, but this is only true for a very limited kind of environment. Intelligence is used to perform well in a wide range of environments.

In the same way, doing nothing is a fixed strategy that only works if something else is achieving your goals when you act too late.

The third kind of expectation requires some explanation. If an intelligent agent is asked whether to choose between A or B within 3 seconds, then the agent should choose A or B in a such way that it could not spend those 3 seconds reflecting on which choice it can imagine picking if there was more time to think, and then come up with a better choice.

The reason for this is obvious: If it can predict about itself making a better choice in the future in the hypothetical scenario where it has a lot of time to think, then it could just use that prediction as a fixed strategy. One thing is to know that a better choice ought to be made, but if a specific choice can be imagined, then the quality of that prediction should be consistent with the actual choice.

Any intelligent agent should make good use of the available time to think. This does not mean that it needs to predict itself. There are situations where one can prove that the decision made is as good as a hypothetical self-prediction, for example by making the best use of time as possible. However, the opposite is not true. If an agent is capable of self-prediction, then it should make at least as a good choice as the predicted choice within the available time.

This is true even if the agent can imagine itself given unlimited time to think. If there exists some piece of knowledge, easily available, that magically reveals which decision it will make in the end, then the agent should stop being clever and just seek out that piece of knowledge.

For example, if the agent is confronted with choices that requires calculations and it can imagine that a calculator would be useful to have in a such situation, then it should just get a calculator, or in the case calculators are hard to bring by demand, be prepared by putting a calculator in its pocket.

This line of reasoning is consistent with Functional Decision Theory (FDT), but it does not tell what underlying principles are used for decisions. Some agents are not capable of self-prediction and are not making optimal decisions according to FDT. Since this is not expected of them, the only requirements to call them intelligent is by making better decisions than random choices or than doing nothing.