where cash rewards are offered to the designers of algorithms that can compress a particular 100-megabyte extract of Wikipedia into ever smaller chunks. However, Hutter is the first to admit that compression alone cannot explain every aspect of intelligence. To fully show off its skills, an AI should also be tested on its ability to use the knowledge it has compressed: it must show it can make smart decisions and plan ahead based on what it knows, he says. You may think it trivial to be able to decide to take an umbrella with you when the sky looks menacing, but it is this kind of pattern recognition that also lets us predict an opponent's move in chess, for instance, and ultimately helped to make us the planet's dominant species.

Designing a practical test that measures these skills mathematically is as hard as it sounds, but Hernández-Orallo and Dowe reckon it is possible to use Kolmogorov complexity to test for decision-making ability and planning, as well as compression.

I got a flavour of how such a test might work by trying out an early version that the pair designed, which they call the "Anytime Intelligence" test. They've already asked both humans and machines to give it a try. Although this "prototype" is a vast simplification and cannot yet be considered a universal test, it serves as a neat demonstration of the principles (*Artificial Intelligence*, DOI: 10.1016/j.artint.2010.09.006).

The Anytime Intelligence test takes the form of a series of interactive tasks displayed on a computer screen. Each task consists of a row of boxes containing symbols. At first there are only three boxes, but the number increases with subsequent tasks. The aim is to gain as many positive "rewards" as possible and to avoid negative ones. I read these instructions beforehand; an AI would be programmed with the rules.

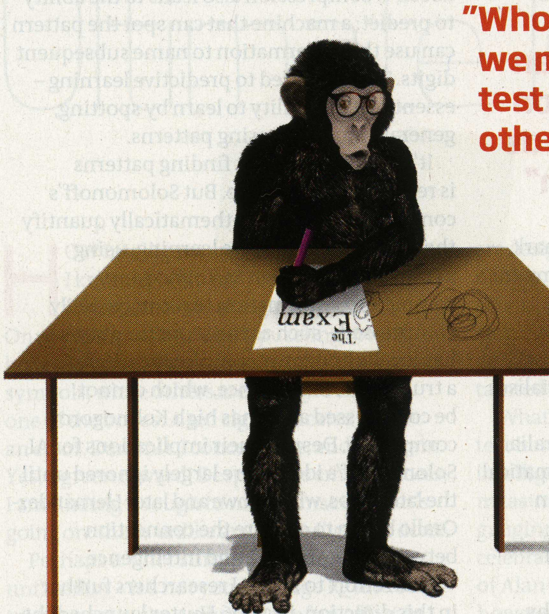
Invisible paths

Once the test began, I could use my mouse to move a symbol that represented "me" between boxes. After each selection, the other symbols in the boxes would then also move, and I would be given feedback that my choice was positive, negative or neutral. This is displayed as green, red and grey signs respectively (see diagram, page 43).

After a while, I began to notice patterns. The key was that moving to certain boxes was not allowed. It turned out an invisible network of "paths" joined some boxes to others. Using these paths to "chase down" one of the two types of symbol would lead to positive rewards, and vice versa for the other symbol.

OK, but how does all that put a figure on intelligence? The patterns of paths and the behaviour of the other symbols within this environment can be expressed as a string of bits, whose Kolmogorov complexity can be estimated. So, your ability to identify the patterns, as gauged by the rewards you win, can be used to calculate a score. In other words, the test is mathematically assessing your ability to spot, compress and then reuse knowledge.

"Who knows what surprises we might find if we could test the intelligence of other creatures fairly?"



What's more, some of the more specific skills belonging to software or humans lend no advantage. Hernández-Orallo and Dowe realised that creatures used to navigating spatial environments, like us, might spot certain patterns more easily than a computer: for example, we might be more inclined to notice paths between physically adjacent boxes. So they made sure that the paths are generated randomly. Conversely, the test is untimed, which means a machine's ability to make rapid calculations is not favoured either.

Hernández-Orallo, Dowe and their team asked 20 people to take 20 variants of the prototype test. They also tested a machine-learning algorithm called Q-learning, which was chosen because it is programmed to learn on the basis of the rewards it is given. They presented the experiment at the Artificial General Intelligence conference in Mountain View, California, in August.

The results revealed far more about the challenges involved in building a universal intelligence test than they do about the intelligence of the participants. The Q-learning algorithm got a slightly better average score than the people. Yet no one would suggest that Q-learning's intelligence is anywhere close to a human's, says Hernández-Orallo. "Q-learning is quite a stupid system," he admits.

So how to make a more effective universal test? One of the first steps will be to make the test respond to individual performance. The prototype does not adapt to the intelligence of the being taking it. It should become harder if someone is acing it, and easier if they are not doing well. This would ensure that smarter participants – such as humans – get given harder tests, and, as a result, the opportunity to shine. This could also prevent boredom if the test is either too easy or too hard – a problem with the test I took that certainly occurred to me as did it.

To allow the test to be taken by animals – not to mention aliens – the interface will have to be redesigned. You might struggle to get a dog to sit down at a screen, and a dolphin can hardly operate a computer mouse. Animal psychologists have wrestled with such problems in the past. "We can compare different species on some basic tasks