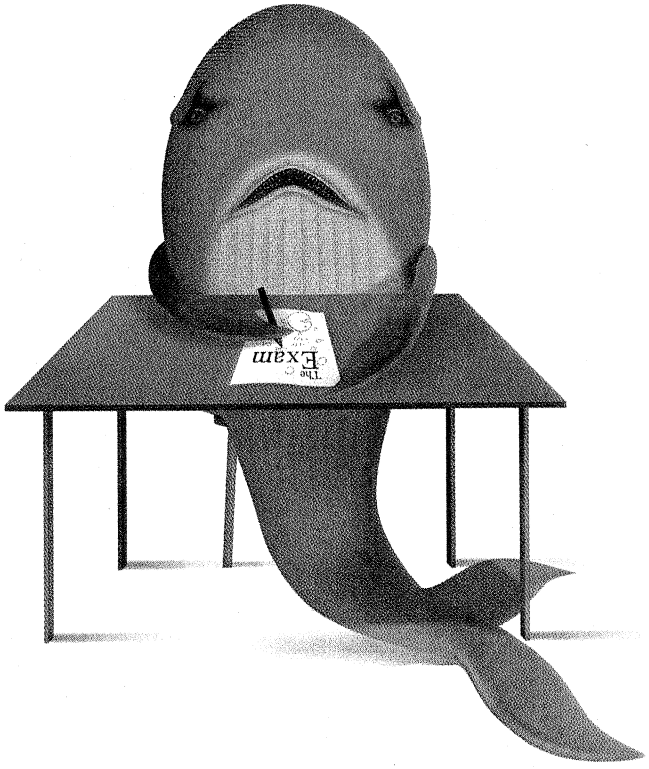
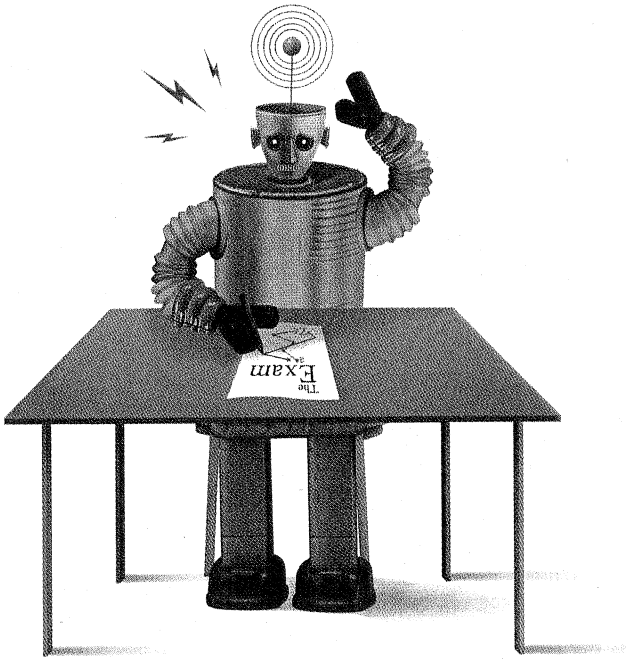




It's time to abandon the human mind as the gold standard of intelligence, says **Celeste Biever**

Ultimate IQ

HOW intelligent are you? I'd like to think I know how smart I am, but the test in front of me is making me reconsider. On my computer screen, a puzzling row of boxes appears: some contain odd-looking symbols, while others are empty. I click on one of the boxes. A red sign indicates I made an error. Dammit. I concentrate, and try again. Yes, a green reward! Despite this small success, I am finding it tough to make sense of what's going on: this is unlike any exam I've ever done.

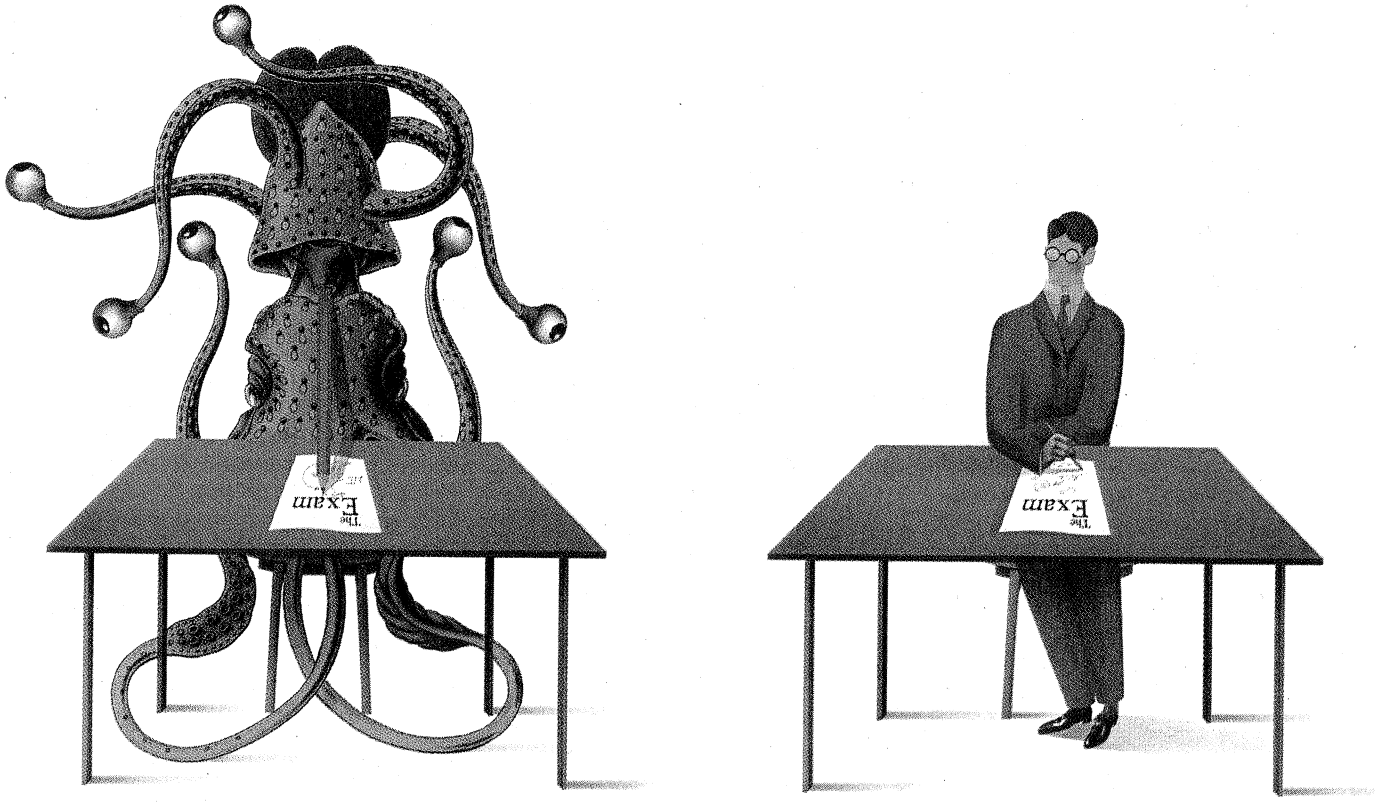
Perhaps it's not surprising that it feels unfamiliar – it's not your average IQ test. I am taking part in the early stages of an effort to develop the first "universal" intelligence test. While traditional IQ and psychometric tests are designed to home in on differences between people, a universal test would rank humans, robots, chimps and perhaps even aliens on a single scale – using a mathematically derived definition of intelligence, rather than one tainted by human bias.

BRETT RYDER

What's the point? The idea for a universal test has emerged from the study of artificial intelligence and a desire for better ways to measure it. Next year, the most famous test for gauging the smarts of machines will be widely celebrated on the 100th anniversary of the birth of Alan Turing, its creator. The Turing test is, however, flawed. To pass it, a machine has to fool a human judge into believing he or she is conversing with another person. But exactly how much smarter are you than the cleverest robot? The test cannot tell you. It also cannot measure intelligence greater than a human's. Machines are getting smarter – possibly smarter than us, soon – so we need a much better way to gauge just how clever they are.

But a universal intelligence test will do more than provide a tool for AI research. Should we encounter an advanced civilisation from another planet, a test based on mathematical principles might tell us what we are dealing with. And here on Earth, it could help us identify life forms that display unfamiliar types of intelligence – who says ours is the only kind? In fact, devising a test free of human bias may be a route to discovering the true nature of intelligence itself. "Just using one species, it is very difficult to be precise about what intelligence is," says José Hernández-Orallo at the University of Valencia in Spain, who is one of the idea's proponents. We have always considered ourselves the gold standard of intelligence, but it's time to give up the notion that our brains are the benchmark.

Testing our own intelligence is easy enough via IQ tests, despite a few recognised flaws. But when it comes to accurately measuring non-humans, these tests are useless. They are not based on a mathematical or even formal



definition of intelligence, and often assume knowledge and skills that are unique to us.

For AI, the best-known performance test is Turing's. However, finding a program that can imitate the abilities of the human mind has proved to be a big challenge. Since 1990, the Loebner prize competition, based on the Turing test, has sparked the creation of a multitude of "chatbots" with fairly impressive social skills, yet most AI researchers don't think these are truly smart. "The Turing test leads to interesting philosophical arguments about intelligence in general and how can we measure it – but it has never been taken seriously as the ultimate goal for AI," says Marcus Hutter at the Australian National University in Canberra.

Instead, myriad specific tests measure "narrow" types of AI – for example, the ability to play chess. IBM's Deep Blue beat Garry Kasparov in 1997, and yet it would be utterly useless at adapting itself to complete a crossword or even to figure out the best way to fold your clothes.

One attempt to encourage the development of AIs with broader intelligence is the General Game Playing competition, held annually at the meeting of the American Association for Artificial Intelligence. Bots are served up a combination of games, from noughts and crosses (tic-tac-toe) to draughts (checkers). They must then devise their own game plans using nothing but a list of rules for those games given to them beforehand.

However, the contest is still asking machines to play at being human. Could there be another, independent, benchmark for their intelligence? If so, we could compare machines with each other much more accurately – as well as with ourselves.

"If we encounter aliens, an intelligence test based on mathematics would tell us what we're dealing with"

Hernández-Orallo decided to devise such a test, along with David Dowe, who specialises in information theory and statistics at Monash University in Melbourne, Australia. For inspiration, they turned to a mathematical definition of intelligence with its roots in the 1960s.

Back then, AI pioneer Ray Solomonoff related intelligence to the ability to summarise or "compress" information by detecting patterns. This skill allows for better problem-solving than using mere trial and error. For example, faced with the sequences 101010101010 or 1234567, a machine or person that realises these can be summarised as "repeat '10' seven times" or "count to 7" is rated as more intelligent than one that doesn't. Compression also leads to the ability to predict: a machine that can spot the pattern can use that information to name subsequent digits. This is related to predictive learning – essentially the ability to learn by spotting, generalising and reusing patterns.

It's nothing new that finding patterns is related to intelligence. But Solomonoff's contribution was to mathematically quantify the process of predictive learning, using a concept now known as Kolmogorov complexity. Information that can be easily compressed – such as the sequences above – has low Kolmogorov complexity, whereas a truly random sequence, which cannot be compressed at all, has high Kolmogorov complexity. Despite their implications for AI, Solomonoff's ideas were largely ignored until the late 1990s, when Dowe and later Hernández-Orallo began to explore the connection between compression and intelligence.

In an effort to spur AI researchers further in this direction, in 2006 Hutter launched the Human Knowledge Compression prize, commonly known as the Hutter prize, ›

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.

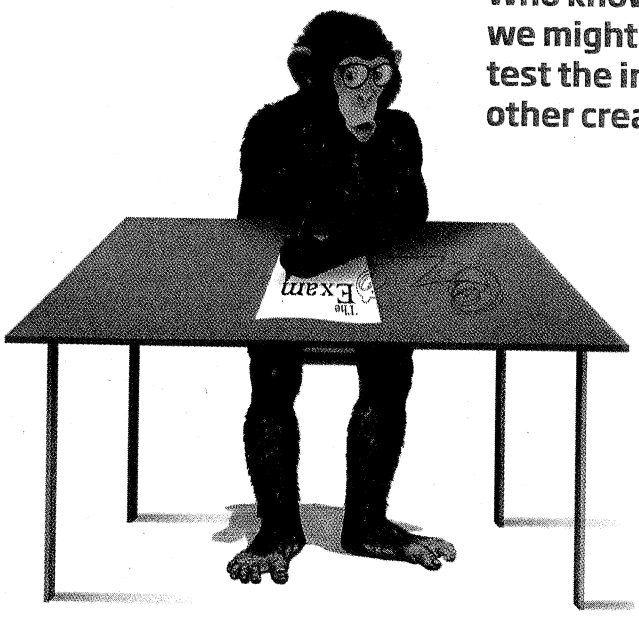
Invisible paths

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.

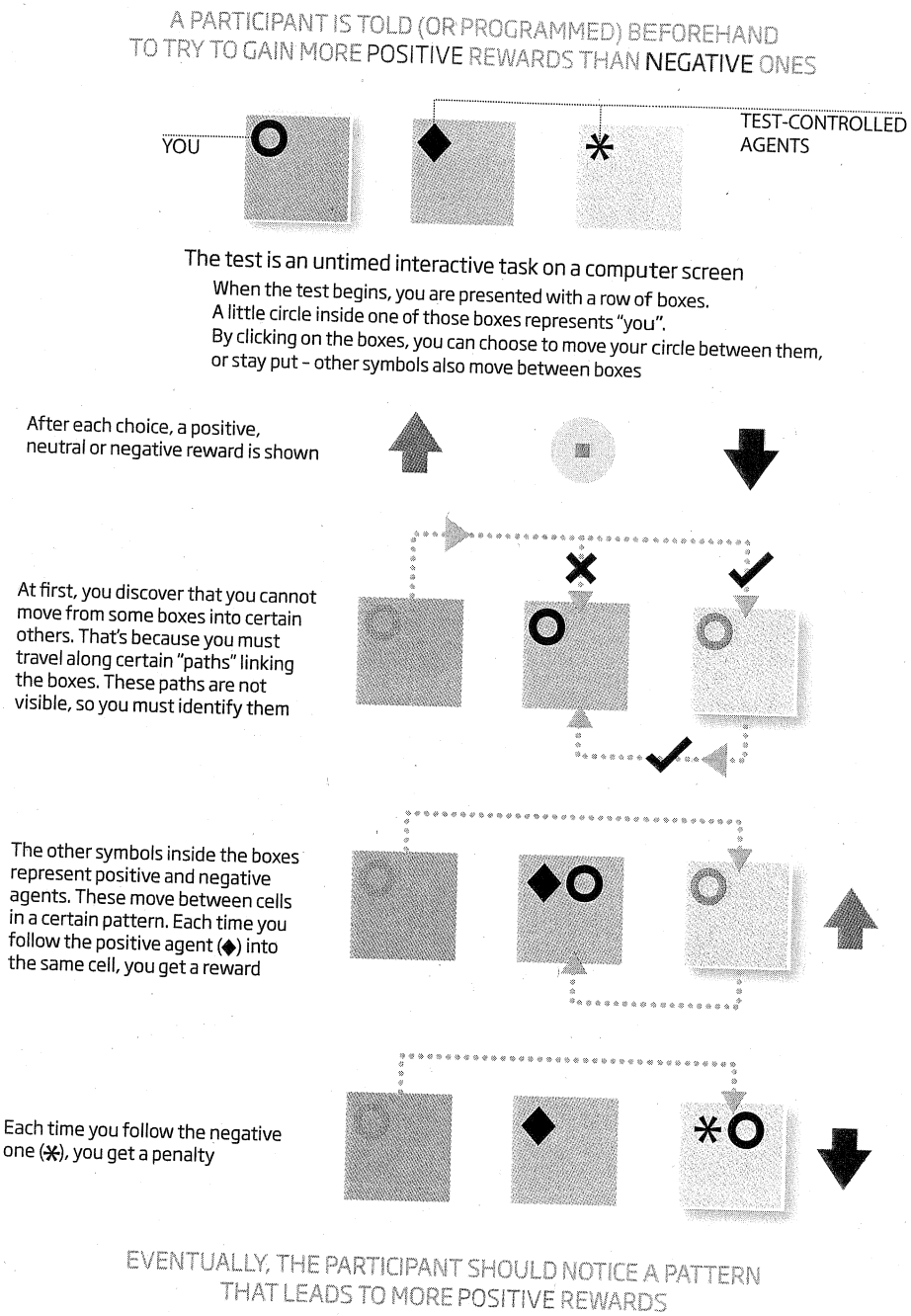
The results revealed far more about the challenges involved in building a universal intelligence test than they do about the intelligence of the participants. The Q-learning algorithm got a slightly better average score than the people. Yet no one would suggest that Q-learning's intelligence is anywhere 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

Spot the pattern

How might a test of universal intelligence work? This test, simplified here, aims to place both humans and machines on the same intelligence scale. It calculates the participant's ability to spot hidden patterns



and compare performance," says Douglas Detterman at Case Western Reserve University in Cleveland, Ohio. "The problem is that in order to show optimum performance, tasks have to be designed for each species. For example, since you cannot give written instructions, how do you make it so each animal approaches the test on an equal footing?

Hernández-Orallo has enlisted experts in animal cognition to help deal with these issues, but a similar problem could even occur with machines: different algorithms might respond differently to the same programmed instructions, which would not necessarily be down to their intelligence.

"I expect it to take quite some research effort until all the kinks are ironed out," says Tom Schaul, an AI researcher at the University of Lugano in Switzerland, who is working on a rival general test for machines. He believes you could design an algorithm that aces the test but performs abysmally on most other tasks. However, he also thinks Hernández-Orallo and Dowe are on the right track. "What I like most is that they succeeded in producing a test that appears to not prefer human agents over artificial ones or vice versa," he says. A formal, testable and completely general measure of intelligence is "crucial for the future of AI", he adds.

If it works, the implications and benefits of a truly universal test would be far-reaching. "Understanding what intelligence is can't be separated from the problem of how intelligence can be measured," says Hernández-Orallo, in the same way that energy, distance and velocity can only be understood through our ability to measure them quantitatively. He says that our current inability to quantify intelligence in a general sense outside of the human species is a major problem, relegating it to a philosophical idea, rather than a scientific one.

Smart slime

There's certainly potential for discovery if we can move beyond our human-centric view of intelligence. Researchers have already found intelligence in unexpected creatures, such as slime moulds that live in dung, which show a surprising ability to navigate mazes. Cephalopods, too – squids, octopuses and the like – have a mental prowess that has only recently been properly appreciated.

Who knows what other surprises we might find on Earth – not to mention in space – if we could design a test to assess other beings fairly? "Exploring intelligence through the special case of human intelligence is seriously mistaken," says Blay Whitby, a philosopher at the University of Sussex in Brighton, UK, specialising in AI. "If we relaxed the requirement that it has to be like us, we might see a lot more intelligence."

Perhaps we would also show a bit more appreciation for the AIs we have already created here on Earth, from the sophisticated search algorithms that let us navigate the web, to the programs that have so much influence over the stock markets. "Once you drop the anthropocentric requirement, AI looks a lot more impressive," Whitby says.

Thinking through all this, I recall my first stab at Hernández-Orallo's test: the truth is I found all those coloured boxes so confusing and frustrating that I simply gave up. I doubt a machine would do the same. Just how flawed human intelligence can be has never felt so apparent. ■

Celeste Biever is *New Scientist*'s deputy news editor