DO WE LIVE IN A COMPUTER SIMULATION?

NICK BOSTROM

SCIENCE has revealed much about the world and our position within it. Generally, the findings have been humbling. The Earth is not the centre of the universe. Our species descended from brutes. We are made of the same stuff as mud. We are moved by neurophysiological signals and subject to a variety of biological, psychological and sociological influences over which we have limited control and little understanding.

One of our remaining sources of pride is technological progress. Like the polyps that over time create coral reefs, the many generations of humans that have come before us have built up a vast technological infrastructure. Our habitat is now largely one of human making. The fact of technological progress is also in a sense humbling. It suggests that the most advanced technology we have today is extremely limited and primitive compared with what our descendants will have.

If we extrapolate these expected technological advances, and think through some of their logical implications, we arrive at another humbling conclusion: the "simulation argument", which has caused some stir since I published it three years ago.

The formal version of the argument requires some probability theory, but the underlying idea can be grasped without mathematics. It starts with the assumption that future civilisations will have enough computing power and programming skills to be able to create what I call "ancestor simulations". These would be detailed simulations of the simulators' predecessors – detailed enough for the simulated minds to be conscious and have the same kinds of experiences we have. Think of an ancestor simulation as a very realistic virtual reality environment, but one where the brains inhabiting the world are themselves part of the simulation.

The simulation argument makes no assumption about how long it will take to develop this capacity. Some futurologists think it will happen within the next 50 years. But even if it takes10 million years, it makes no difference to the argument.

Let me state what the conclusion of the argument is. The conclusion is that at least one of the following three propositions must be true:

1 Almost all civilisations at our level of development become extinct before becoming technologically mature.
2 The fraction of technologically mature civilisations that are interested in creating ancestor simulations is almost zero.
3 You are almost certainly living in a computer simulation.

How do we reach this conclusion? Suppose first that the first proposition is false. Then a significant fraction of civilisations at our level of development eventually become technologically mature. Suppose, too, that the second proposition is false. Then a significant fraction of these civilisa-

tions run ancestor simulations. Therefore, if both one and two are false, there will be simulated minds like ours.

If we work out the numbers, we find that there would be vastly many more simulated minds than non-simulated minds. We assume that technologically mature civilisations would have access to enormous amounts of computing power.

So enormous, in fact, that by devoting even a tiny fraction to ancestor simulations, they would be able to implement billions of simulations, each containing as many people as have ever existed. In other words, almost all minds like yours would be simulated. Therefore, by a very weak principle of indifference, you would have to assume that you are probably one of these simulated minds rather than one of the ones that are not simulated.

Hence, if you think that propositions one and two are both false, you should accept the third. It is not coherent to reject all three.

It should be emphasised that the simulation argument does not show that you are living in a simulation. The conclusion is simply that at least one of the three propositions is true. It does not tell us which one.

In reality, we don't have much specific information to tell us which of the three propositions might be true. In this situation, it might be reasonable to distribute our credence roughly evenly between them.

Let us consider the options in a little more detail. Proposition one is straightforward. For example, maybe there is some technology that every advanced civilisation eventually develops and which then destroys them. Let us hope this is not the case. Proposition two requires that there is a strong convergence among all advanced civilisations, such that almost none of them are interested in running ancestor simulations. One can imagine various reasons that may lead civilisations to make this choice. Yet for proposition two to be true, virtually all civilisations would have to refrain. If this were true, it would be an interesting constraint on the future evolution of intelligent life.

The third possibility is philosophically the most intriguing. If it is correct, you are almost certainly living in a computer simulation that was created by some advanced civilisation. What Copernicus and Darwin and latter-day scientists have been discovering are the laws and workings of the simulated reality. These laws might or might not be identical to those operating at the more fundamental level of reality where the computer that is running our simulation exists (which, of course, may itself be a simulation). In a way, our place in the world would be even humbler than we thought.

What kind of implications would this have? How should it change the way you live your life?

Your first reaction might think that if three is true, then all bets are off and you would go crazy. To reason thus would be an error. Even if we are in a simulation, the best methods of predicting what will happen next are still the familiar ones – extrapolation of past trends, scientific modelling and common sense. To a first approximation, if you thought you were in a simulation, you should get on with your life in much the same way as if you were convinced that you were leading a non-simulated life at the "bottom" level of reality.

If we are in a simulation, could ever know for certain? If the simulators don't want us to find out, we probably never will. But if they choose to reveal themselves, they could certainly do so. Another event that would let us conclude with a high degree of confidence that we are in a simulation is if we ever reach a point when we are about to switch on our own ancestor simulations. That would be very strong evidence against the first two propositions, leaving us only with the third.

Nick Bostrom is the director of the Future of Humanity Institute at the University of Oxford