The end of the world and life in a computer simulation

PHIL 20229
Jeff Speaks

February 7, 2008

1 The doomsday argument

Leslie begins his exposition of the argument with an example:

> **Prima facie, we should prefer theories which make our observations fairly much to be expected, rather than highly extraordinary. Waking up in the night, you form two theories. Each has a half-chance of being right, you estimate. The first, that you left the back door open, gives the chances as 10 per cent that the neighbour's cat is in your bedroom. The second, that you shut the door, puts those chances at 0.01 per cent. You switch on the light and see the cat. You should now much prefer the first theory.**

This reasoning is, fairly clearly, a kind of reasoning upon which we rely all of the time. It might be summed up like this: if we have two theories, and the first makes a certain event much more likely than the other, and we observe that event, that should lead us to favor the first theory. Note that this doesn't mean that we should always think that the first theory is true; rather, what it means is that, whatever our initial estimate of the probability of the first theory, we should increase that estimate upon observing the event in question.

Now consider the application of that line of reasoning to the examples of balls shot at random from a lottery machine. Suppose that you know that the balls in the machine are numbered sequentially (with no repeats) beginning with 1, but that you don't know how many balls there are in the machine. Now we start the machine, and a ball comes out with '3' on it. You're now asked: do you think that it is more likely that the machine has 10 balls in it, or 10,000 balls in it? The line of reasoning sketched above seems to favor the hypothesis that there are just 10 balls in the machine. (Note that, whichever hypothesis you endorse, you could be wrong; this is not a form of reasoning which delivers results guaranteed to be correct. The question is just which of these hypotheses is most likely, given the evidence.)

In fact, we can do better than just saying that in such cases you should raise the probability you assign to one theory. We can, using a widely accepted rule of reasoning called 'Bayes' theorem', say how much you should raise your probability assignment. (One reason why this theorem is widely accepted is that following it enables one to avoid 'Dutch book' arguments.)

To arrive at Bayes' theorem, we can begin with the definition of what is called 'conditional probability': the probability of one claim, given that another is true. In particular, for arbitrary claims a and b, we can say that

$$P(a|b) = \frac{P(a\&b)}{P(b)}$$

In other words, the probability of a given b is the chance that a and b are both true, divided by the chances that b is true. For example, let a = 'Obama wins', and let b = 'a man wins.' Suppose that each of Obama, Hilary, and McCain have a 1/3 chance of winning. Then the conditional probability is that Obama wins, given that a man wins, is 1/2. Intuitively, if you found out only that a man would win, you should then (given the initial probability assignments) think that there is a 0.5 probability that Obama will win.

Using this definition of conditional probability, we can then argue as follows, assuming that $P(b) \neq 0$:

1. $P(a\|b) = \frac{P(a\&b)}{P(b)}$	def. of conditional probability
2. $P(b\|a) = \frac{P(a\&b)}{P(a)}$	def. of conditional probability
3. $P(a\|b) * P(b) = P(a\&b)$	(1), multiplication by =';s
4. $P(a\&b) = P(b\|a) * P(a)$	(2), multiplication by ='s
5. $P(a\|b) * P(b) = P(b\|a) * P(a)$	(3),(4)
C. $P(a\|b) = \frac{P(b\|a)*P(a)}{P(b)}$	(5), division by ='s

This conclusion is Bayes' theorem. Often, what we want to know is, intuitively, the probability of some hypothesis 'h' given some evidence 'e'; then we would write the theorem as:

$$P(h|e) = \frac{P(h)*P(e|h)}{P(e)}$$

Consider what this would say about the example of the lottery machine. Suppose for simplicity that you know going in that there are only two options, which are equally likely to be correct: that there are 10 balls in the machine, and that there are 10,000. Let e be the evidence that the first ball to come out is #3, and let h be the hypothesis that there are 10 balls in the machine. Then we might say:

$$P(h) = 0.5$$
$$P(e|h) = 0.1$$
$$P(e) = 0.5(0.1 + 0.0001) = 0.05005$$

Then we find, via Bayes' theorem, that $P(h|e) = \frac{0.5*0.1}{0.05005} = 0.999$. So, on the basis of the evidence that the first ball to come out was #3, you should revise your confidence in the 10-ball hypothesis from 50% to 99.9% certainty.

Bayes' theorem can be restated in the following way:

$$P(h|e) = \frac{P(h)*P(e|h)}{P(h)*P(e|h)+P(not-h)*P(e|not-h)}$$

(Don't worry about the proof of this being equivalent to Bayes' theorem.) This way of putting the theorem will be important for understanding Leslie's argument.

1.2 Anthropic reasoning

Next, Leslie introduces what he calls 'the anthropic principle' or 'anthropic reasoning.' He never defines this sort of reasoning, but for our purposes we can think of it as just pointing out that we can treat our observation about when and where we exist as an observation which, like any other, might be used to adjust the probabilities which we assign to various hypotheses.

Consider, in particular, the observation that you are (say) the 50 billionth person (in order of birth) to exist. Let e be the claim that you were among the first 50 billion persons born. Now consider the following two hypotheses:

> Doom Soon: The human race will go extinct by 2150, with the total humans born by the time of such extinction being 500 billion.
>
> Doom Delayed: The human race will go on for several thousand centuries, with the total humans born before the race goes extinct being 50 thousand billion.

Suppose that we think, before considering where we are in the order of human births, that there is a 1% chance of Doom Soon, and a 99% chance of Doom Delayed.

Given these initial probabilities, how should we calculate P(Doom Soon|e)? (See Leslie, pp. 200-203.) What do we have to assume about our place in the birth order of human history for this argument to work?

2 Responses to the argument

Objection 1: Couldn't someone in ancient Rome have used this reasoning to show that the end the world would come before 500 AD? And wouldn't they have been wrong? So mustn't there also be something wrong with our using this reasoning?

Objection 2: Our place in the birth order was not, in any good sense, randomly selected. But random selection is required for the analogy with the lottery machine to be a good one. If, for example, the machine was rigged to favor lower numbered balls, this would clearly affect our response to the above example.

Reply to objection 2: There's no reason to believe that there's any correlate of 'rigging' in the present case; there's no reason to believe that my rank in the birth order of human history is anything other than typical.

Objection 3: The argument neglects an important piece of evidence: not just that you exist now, but that you exist at all. That you exist is much more likely on Doom Delayed than on Doom Soon; for, of the 50 thousand billion people who would exist if Doom Delayed is true, only 1% of them would be born if Doom Soon is true. This counterbalances the force of the evidence that you are among the first 50 billion humans born. (See the Greenberg article on the course web site.)

Reply to objection 3: But you had to have been born; otherwise you wouldn't be here to consider the argument!

Is this a good reply to the objection? An analogy with the example of the firing squad.

Another reply to the objection: Remember again the case of the lottery. Surely the fact that ball #3 comes out first *does* give us good evidence that there are 10 rather than 1000 balls in the machine. But you could reply to that argument that the fact that a given ball comes out is much *more* likely on the 1000 ball hypothesis, since on the 10 ball hypothesis 990 of those balls would not even have been in the machine, and hence would not even have the chance to come out! But surely there's something wrong with this reply since, after all, the fact that ball #3 comes out first *does* give us good evidence that there are 10 rather than 1000 balls in the machine.

A reply to the reply: This is why the two cases are different. The lottery balls are ordered by the numbers written on them; this rules out a scenario in which, for example, balls 570-579 and no others are in the machine. But the doomsday argument assumes that our place in the birth order is random, which is in effect to assume that there is no intrinsic ordering of human births. Suppose that we knew that there were either 10 balls or 1000 in the machine, but did not know, given that the first scenario was true, whether the balls in the machine would be 1-10; 570-579; 56-59, 476, 999, and 1-5; or any other random collection of 10 numbers. Now the #3 ball shoots out. Does that give you any evidence that there are 10 rather than 1000 balls in the machine?

3 Are we living in a computer simulation?

The doomsday line of argument can be extended in a surprising way, if we make a few further assumptions. Consider the following line of argument: We could meet Martians, or creatures made of very different physical stuff than us, who were nevertheless conscious and had mental lives which were as rich and varied as our own. So it seems that having a mental life like ours is not dependent on being human in particular, but rather on having a certain kind of complex organization. If so, then it seems likely that computers will eventually reach that level of complexity, and become conscious. Indeed, with sufficient increases in computing power, it seems likely that computers will be able to run simulations of human life which contain many, many conscious beings. With continuing increases in computing power, there will come a time at which a very high percentage of all beings in history who have had a consciousness like ours will have been computer simulations.

If this is so, isn't it reasonable for you to now infer that you are living in a computer simulation (or, perhaps, that the human race will end relatively soon, before such computing advances are made)?

Is this argument related to the Doomsday argument? If so, how?