**A note on the Simulation Argument**

**Peter J. Lewis**

[**plewis@miami.edu**](mailto:plewis@miami.edu)

The point of this note is to compare the Doomsday Argument to the Simulation Argument. The latter, I maintain, is a better argument than the former.

As has been shown elsewhere, the Doomsday Argument fails; learning our birth rank among humans does not confirm a short future for the human race over a longer one (Lewis 2010, following Pisaturo 2009). This is so even if learning our birth rank confirms a small total human population over a larger one. This might sound paradoxical, but it is not; even though your evidence makes it less likely that the population is large, if it *is* large, you must be at the beginning of it, and hence the future is long. Call this the relative location effect.

The diagram in fig. 1 makes this clear. Suppose for simplicity that humanity consists either of one person (hypothesis H1­), two people in succession (H2), or three people in succession (H3). You are either the first person (hypothesis E1), the second (E2) or the third (E3). Initially suppose that your credences are distributed as on the left. Then if you learn that your birth rank is low (i.e. you learn E­1), this tells you that P(E2) = 0 and P(E3) = 0. Renormalizing your other credences (i.e. applying Bayes’ theorem) yields the credences on the right. Note that H1 has been confirmed; your credence in H1 goes from 1/6 to 1/3. That is, discovering that your birth rank is low increases your credence in a small overall population.[[1]](#footnote-1) But now consider hypothesis D, the hypothesis that you are the last person, represented by the diagonal in the diagram. Your credence in D has decreased from 1/2 to 1/3, and your credence in ~D has increased from 1/2 to 2/3. That is, discovering that your birth rank is low decreases your credence in a short future, and increases your credence in a longer future.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| H1 | 1/6 |  |  | H1 | 1/3 |  |  |
| H2 | 1/6 | 1/6 |  | H2 | 1/3 | 0 |  |
| H3 | 1/6 | 1/6 | 1/6 | H3 | 1/3 | 0 | 0 |
|  | E1 | E2 | E3 |  | E1 | E2 | E3 |

Fig. 1: Doomsday Argument

These results generalize straightforwardly to any finite number of hypotheses about the total human population, provided your credence is initially uniformly distributed among your possible locations.[[2]](#footnote-2) Suppose you consider *n* population hypotheses, where H*i* says that *i* people are created, and *n* corresponding location hypotheses, where E*j* says that your birth rank is *j*. Then each possible location H*i* & E*j* (for *j* ≤ *i*) has initial credence P(H*i* & E*j*) = 1/*s*, where *s* = 1 + 2 + … + *n*. Initially, then, P(H1) = 1/*s* and P(E1) = *n*/*s*. After learning E1, your credence in H1 shifts to P′(H1) = P(H1)/P(E1) = 1/*n*. Since 1/*n* > 1/*s*, the small population hypothesis H1 is confirmed. Your initial credence that you are the last person is P(D) = *n*/*s*. After learning E1, your credence in D shifts to P′(D) = P′(H1) = 1/*n*. Since *n*/*s* > 1/*n* for *n* > 1, the short future hypothesis D is disconfirmed.

Now compare the Simulation Argument (Bostrom 2003; 2009). It is possible that we will develop technology that allows us to produce “virtual” conscious being by means of computer simulation; in fact, it is possible that we *are* such simulations. Furthermore, it is possible for simulations to produce simulations of their own; so there could be a two-level hierarchy of unsimulated beings producing simulations, or a three-level hierarchy of unsimulated beings producing simulations producing simulations, and so on. Indeed, thinking about such hierarchies might incline you to think that simulated conscious beings may well outnumber unsimulated ones, making it more likely than not that you are a simulation.

But it is worth noticing that, as yet, we have not simulated any conscious beings. Is this relevant to my credence that I am a simulation? Surely it is, since the fact that we produce no simulations is more likely under a short-hierarchy hypothesis rather than a long-hierarchy hypothesis, and most likely of all under a one-level hypothesis in which there are no simulated beings (Richmond 2008). Hence it looks at first glance like the evidence that we produce no simulations should decrease our credence that there are any simulations, and hence decrease our credence that we are ourselves simulations. I do not dispute the first step in this argument, but the relative location effect ought to make us wary of the second.

So consider the diagram in fig. 2. Suppose for simplicity that there is either a one-level hierarchy (hypothesis H1), a two-level hierarchy (H2), or a three-level hierarchy (H3). You are either unsimulated (hypothesis E1), simulated by unsimulated beings (E2), or simulated by simulated beings (E3). Suppose that your credences are initially as on the left, and you learn that you do not (now) produce simulations. What you learn is that you are in the diagonal D of the diagram; the off-diagonal elements are zero. Renormalizing your other credences (i.e. applying Bayes’ theorem) yields the credences on the right. Note that your credence in H1 goes from 1/6 to 1/3; the evidence that you don’t simulate confirms the hypothesis that there are no simulated beings (Richmond’s point). But note also that your credence that you are a simulated being (i.e. your credence in ~E1) goes up from 1/2 to 2/3.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| H1 | 1/6 |  |  | H1 | 1/3 |  |  |
| H2 | 1/6 | 1/6 |  | H2 | 0 | 1/3 |  |
| H3 | 1/6 | 1/6 | 1/6 | H3 | 0 | 0 | 1/3 |
|  | E1 | E2 | E3 |  | E1 | E2 | E3 |

Fig. 2: Simulation Argument

In other words, the evidence that we don’t simulate confirms both the hypothesis that there are no simulated beings and the hypothesis that you are a simulated being. As with the Doomsday result above, this might appear paradoxical. But it is not, because *if* there are simulated beings, your evidence confirms that you are one of them.

Again, this generalizes straightforwardly to any finite number of hypotheses about the size of the simulation hierarchy, provided your credence is initially uniformly distributed over your possible locations.[[3]](#footnote-3) Suppose there are *n* possible simulation hypotheses H*i* (that there is now an *i*-level simulation hierarchy) and *n* corresponding location hypotheses E*j* (that you are in the *j*th level). Let D be the hypothesis that you do not simulate, represented by the diagonal elements *i* = *j*. As before, each possible location H*i* & E*j* (for *j* ≤ *i*) has initial credence P(H*i* & E*j*) = 1/*s*, where *s* = 1 + 2 + … + *n*. Initially, then, P(H1) = 1/*s* and P(D) = P(E1) = *n*/*s*. After learning D, your credence in H1 shifts to P′(H1) = P(H1)/P(D) = 1/*n*. Since 1/*n* > 1/*s*, the hypothesis H1 that there are no simulated beings is confirmed. Your initial credence that you are not simulated is P(E1) = *n*/*s*. After learning D, your credence in E1 shifts to P′(E1) = P′(H1) = 1/*n*. Since *n*/*s* > 1/*n* for *n* > 1, the hypothesis E1 that you are not simulated is disconfirmed.

I conclude that the Simulation Argument is on firmer ground than the Doomsday Argument, since the relative location effect supports the former but undermines the latter. This seems exactly right. The Doomsday Argument looks like a fallacy; surely my birth rank is irrelevant to the future of humanity. But the Simulation Argument looks like a familiar form of skepticism, requiring a philosophical response (e.g. Chalmers 2005) rather than a logical refutation.

**Acknowledgments**

I would like to thank Ronald Pisaturo for helpful comments on an earlier draft.

**References**

Bostrom, Nick (2003), “Are we living in a computer simulation?”, *Philosophical Quarterly* 53: 243–255.

——— (2009), “The Simulation Argument: some explanations”, *Analysis* 69: 458-461.

Chalmers, David J. (2005), “The Matrix as metaphysics”, in Christopher Grau (ed.), *Philosophers Explore the Matrix*. Oxford: Oxford University Press, 132–176.

Dieks, Dennis (2007), “Reasoning about the future: Doom and Beauty”, *Synthese* 156: 427–439.

Lewis, Peter J. (2010), “A note on the Doomsday Argument”, *Analysis* 70: 27–30.

Pisaturo, Ronald (2009), “Past longevity as evidence for the future”, *Philosophy of Science* 76: 73–100.

Richmond, Alasdair M. (2008), “Doomsday, Bishop Ussher and simulated worlds”, *Ratio* 21: 201–217.

1. It is perhaps misleading to say that H1 is confirmed, since one can plausibly argue that this probability shift simply restores your credence in H1 to its value before you started thinking about your birth rank (Dieks 2007; Lewis 2010). [↑](#footnote-ref-1)
2. This is essentially the “thirder” distribution from the Sleeping Beauty literature. If one adopts instead the “halfer” distribution, so that the H*i* are initially equiprobable, the situation is more ambiguous; for *n* ≤ 6 the evidence of your birth rank confirms a long future, and for *n* > 6 it confirms a short future. I note here only that the “thirder” distribution is plausible. [↑](#footnote-ref-2)
3. Similar considerations to those in the previous footnote apply to alternative prior probability distributions. [↑](#footnote-ref-3)