The Simulation Argument FAQ

(2025) Nick Bostrom
version 2.0
[simulation-argument.com]

1. What is the simulation argument?

The simulation argument purports to show that *at least one* of the following propositions is true: (1) the human species is very likely to go extinct before reaching a "posthuman" stage; (2) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof); (3) we are almost certainly living in a computer simulation. It follows that the belief that there is a significant chance that we will one day become posthumans who run ancestor-simulations is false, unless we are currently living in a simulation. A number of other consequences of this result are also discussed.

The simulation argument was set forth in a paper published in 2003. A draft had been circulated for a couple of years prior. The argument has attracted a considerable amount of attention, among scientists and philosophers as well as in the media.

References:

Bostrom, N. (2003): "Are You Living in a Computer Simulation?". *Philosophical Quarterly*, Vol. 53, No. 211, pp. 243-255.
Bostrom, N. & Kulczycki, M. (2011): "A Patch for the Simulation Argument". *Analysis*, Vol. 71, No. 1, pp. 54-61.

2. Do you really believe that we are in a computer simulation?

I believe that the simulation argument is basically sound. The argument shows only that at least one of three possibilities obtains, but it does not tell us which one(s). One could accept the simulation *argument* while rejecting the simulation *hypothesis* (i.e. that we are in a simulation).

I would assign a "substantial probability" to the simulation hypothesis. I tend to refrain from providing a specific number. (This is for various reasons, including that it could convey a false sense of precision.) I don't think we have very strong evidence for or against any of the three disjuncts (1)-(3), so it makes sense to allocate our credence between them in a way that gives each of them some non-negligible probability.

I note that people who hear about the simulation argument often react by saying, "Yes, I accept the argument, and it is obvious that it is alternative *n* that obtains." But different people pick a different *n*. Some think it obvious that (1) is true, others that (2) is true, yet others that (3) is true. Many if not all of these people are overconfident.

3. Is the simulation argument a variant of Descartes' daemon or the brain-in-a-vat argument?

No, the simulation argument is fundamentally different from these traditional philosophical arguments (as explained in my reply to Brian Weatherson).

The purpose of the simulation argument is different: not to set up a skeptical problem as a challenge to epistemological theories and common sense, but rather to argue that we have interesting empirical reasons to believe that a certain disjunctive claim about the world is true (i.e., (1) ∨ (2) ∨ (3)). The simulation argument relies crucially on non-obvious empirical premises about future technological abilities. Moreover, the conclusion of the simulation argument is not simply that we cannot be certain that we are not living in a simulation. If we knew that f_{SIM} (the fraction of all human-like beings who are simulated) was very small but non-zero, we might not be able to be completely certain that we are not in a simulation; but that would not be a very interesting contention. (If we think that somewhere in our infinite universe there are a few "envatted brains", then maybe we shouldn't assign a strictly zero credence to us being envatted brains either; but so long as we thought that the proportion of brains in vats to brains in crania was small enough, we would have no ground for seriously doubting that we are not brains in vats, at least if we lacked specific evidence to the contrary.)

The simulation argument is also different from ordinary brain-in-a-vat arguments in that it doesn't begin from a starting point of doubt and then ask for some compelling reason for canceling that doubt. Rather, it begins from the starting point that things are the way we believe they are, and then, while granting us that we might be justified in assigning a high initial credence to these beliefs, nevertheless tries to show that we have specific empirically-grounded reasons for revising these initial beliefs in a certain way—not so as to make us generally agnostic about the existence of an external world, but to accept the disjunctive conclusion. Thus, the simulation argument is not best thought of as a skeptical argument that would have us be more agnostic, but rather as an argument that would have us increase our credence in one particular disjunction (and decrease our credence in its negation). It aims to tell us something about the world rather than to advise us that we know less about the world than we thought we did.

References:

Bostrom, N. (2005): "The Simulation Argument: Reply to Brian Weatherson". *Philosophical Quarterly*, Vol. 55, No. 218, pp. 90-97.

4. If we are in a simulation, doesn't that undermine the reasoning in the simulation argument?

The simulation argument relies on an assumption about the technological capabilities of a mature civilization. Our evidence for this assumption is empirical: it is based on our best theories about the physical limits of computation and the kinds of constructions that could be built with advanced molecular manufacturing techniques etc., and our confidence in these theories rests ultimately on observations of the world we see around us. But if we are in a simulation, then how could we trust these observations? Might they not inform us only about the simulated reality and not about any hypothesized underlying layer of reality in which the simulation is being run? And if so, does this not undermine the simulation argument by casting doubt on its empirical assumptions?

If we reflect more carefully on this objection, however, we see that it fails. The claim that we cannot have any information about the underlying reality if we are in a simulation is false. In a simulation, we can certainly know the following two *conditional* claims:

> A. If we are in a simulation, then the underlying reality is such as to permit simulations, it contains at least one such simulation, and (3) is true.
>
> B. If we are not in a simulation, then the empirical evidence noted in the simulation argument is veridical taken at face value, suggesting that a technologically mature civilization would have the ability to create vast numbers of simulations. We are then back to the reasoning in the simulation argument, from which we infer that there is a high probability that at least one of the disjuncts (1)-(3) is true.

Since we either are, or are not, in a simulation, we can conclude that the disjunct (1) ∨ (2) ∨ (3) is true.

References:

Besnard, F. (2004): "Refutations of the Simulation Argument". *Manuscript*. [Besnard presents a version of this objection; it has also come up in more recent works by others.]

5. "I can see glitches in the Matrix!"

I am very skeptical of such claims. We should expect occasionally to hear this kind of report even if we are not in a simulation. Even if we *are* in a simulation, the most plausible explanation for such reports is not that they result from any real "glitch" but rather that they originate in the

ordinary frailties of the human mind—hallucinations, psychiatric problems, visual illusions, self-deception, misremembering, misinterpretation, fraud, and so forth.

It seems likely that the hypothetical simulators, who would evidently have to be technologically extremely advanced to create simulations with conscious humanlike participants, would also have the ability to prevent these simulated creatures from noticing anomalies in the simulation. This could be done by avoiding anomalies altogether, or preventing them from having noticeable macroscopic ramifications, or by retrospectively editing the brain states of observers who had happened to witness something suspicious. If the simulators don't want us to know that we are simulated, they could easily prevent us from finding out. Consider that even our own humble brains—unaided by technology—usually manage to prevent us from realizing when we are dreaming at night, even though the typical dream is teeming with the most fantastic anomalies.

6. Isn't it computationally infeasible to simulate an entire universe?

It may well be. However, as pointed out in the original paper, that is not required. The simulation argument does not envisage a universe-wide simulation where every atom and every quark is continuously simulated in perfect detail.

Instead, only enough needs to be included in the simulation to make it appear real to the observers inside. This allows many details to be omitted, such as objects that are very small or very far away. Many of the remaining details could be filled in only when somebody is looking at them or performing relevant experiments. Graphics engines typically render only that which is seen by some player character; and some modern computer games use procedural generation, which creates world details as needed depending on where the player goes, thus removing any limits to the size of the virtual world that can be explored.

We should imagine such techniques being perfected to an extreme degree by superintelligent AI, to achieve the required degrees of realism at the lowest possible computational cost. Superintelligent processing might also be used during simulation runs to invent and render details as required, and to patch up any inconsistencies.

The eventual feasibility of sufficiently realistic-appearing simulations may be made more intuitively plausible if we consider: (a) how much computer graphics and virtual reality has advanced in just a few decades; (b) that our own humble biological brains are evidently capable of generating fantasy that to the observer appears quite convincingly realistic—despite the dreamer (or hallucinator) having plenty of experience of both dreaming and being awake, which is in contrast to the case of a simulated person who may never have experienced anything other than their simulated world; (d) that if some bug or noticeable discrepancy were to occur, it could

be patched up with some retrospective brain editing or by re-running the simulation from a save point; (e) that the simulators would presumably be a technologically mature superintelligent civilization that would have had eons to perfect their techniques; and (f) that since the observer brains are part of the simulation, simulators would have easy digital access to all the sensory neurons and all internal mental states (which might allow, for example, impending shifts in gaze and attention to be anticipated).

Critiques based on the assumption that a simulation would have to be fully comprehensive (e.g. Vazza (2025)) thus miss the point. We may also note that even if fully comprehensive detailed simulations *are* possible—which is not inconceivable, since the physics in the basement universe might allow for vastly more powerful computers than does the physics in our observed universe—it would still be unlikely that we are in a fully comprehensive simulation, since simulators could run vastly more simplified simulations than fully comprehensive ones for a given computational budget.

A related objection is that the computational cost of simulation would increase exponentially over time, as simulated civilizations develop their own simulations—in which civilizations may in turn develop *their* own simulations, and so forth. However, simulators could avoid this by stepping in to prevent simulated civilizations from using excessive amounts of computing power, or by ending them shortly after they begin to consume excessive resources (which wouldn't necessarily entail ending the individuals contained in the simulation). One consequence of this may be that—unless the basement universe allows for infinite computations—most civilizations would be leaf nodes of the simulation tree: they are themselves simulated but they will never run genuine simulations of their own.

References:
Vazza, F. (2025): "Astrophysical constraints on the simulation hypothesis for this Universe: why it is (nearly) impossible that we live in a simulation". *Manuscript*

7. What happens to the argument if the world is infinite?

This is an interesting issue, which is deliberately set aside in the original paper. Certainly some modifications are necessary once we admit the possibility of an infinite universe that may contain infinitely many simulated and non-simulated people. In this case, the ratio of simulated people to the total number of people is not defined.

To deal with these infinite cases, we need to do something like thinking in terms of densities rather than total populations. A suitable density-measure can be finite even if the total population is infinite. It is important to note that we need to use some kind of density-measure of

observation types quite independently of the simulation argument. In a "Big World" cosmology, all possible human observations are in fact made by somebody somewhere. (Our world may well be a Big World, so this is not a far-fetched possibility.) To be able to derive any observational consequences from our scientific theories in a Big World, we need to be able to say that certain types of observations are more typical than others. (See my paper "Self-Locating Belief in Big Worlds" for more details on this.)

The most straightforward way of making this notion precise in an infinite universe is via the idea of limit density. Start by picking an arbitrary spacetime point. Then consider a hypersphere centered on that point with radius R. Let f(A) be the fraction of all observations that are of kind A that take place within this hypersphere. Then expand the sphere. Let the typicality of type-A observations be the limit of f(A) as $R \to \infty$.

To apply this idea to the case where we might be living in a simulation, we can use a similar rule, except that we define the seed point to be the location at the bottom level of reality where the computer is located that is ultimately running our simulation, and we do the hypersphere expansion at that level of reality. (One could modify this rule so that it would work also in the case where it is possible that there is no lowest level of reality but rather an infinite regress of simulations within simulations "all the way down".)

References:
Bostrom, N. (2002): "Self-Locating Belief in Big Worlds: Cosmology's Missing Link to Observation". *Journal of Philosophy*, Vol. 99, No. 12, pp. 607-623.

8. Couldn't we simply be in a very early generation, so that all the simulated creatures that will one day be created don't yet exist?

Indeed we could, but the question is what probability we should assign to this possibility. If we assume that the total number of simulated people with experiences like ours that will eventually have lived is vastly greater than the number of non-simulated people with such experiences that will ever have lived, then we should believe with very high probability that we are among the simulated majority rather than the non-simulated minority. None of all these simulated and non-simulated people have any way of telling which generation they are in. If they all guess that they are in a very early generation, then almost everybody will guess wrong. If they all guess that they are not in a very early generation, then almost everybody will guess right.

The part of the argument most directly pertinent to this point is the "Bland Indifference Principle", defended briefly in section V of the original paper. For a defense of a much stronger principle

from which the Bland Indifference Principle can be derived as a trivial special case, see my book *Anthropic Bias*.

(Incidentally, this objection became the focal point during the second half of a long conversation on the popular Joe Rogan Experience podcast—Joe's notorious "fourth option".)

References:
Bostrom, N.: *Anthropic Bias: Observation Selection Effects in Science and Philosophy* (Routledge, 2002).
Bostrom, N. (2005): "The Simulation Argument: Reply to Brian Weatherson". *Philosophical Quarterly*, Vol. 55, No. 218.
Rogan, J. (2019): *The Joe Rogan Experience # 1350 - Nick Bostrom*. 12 September 2019. [The conversation about the simulation argument starts around 1:17:00.]

9. How has the simulation argument affected how you live?

It has clearly influenced my intellectual work, since the constraint it identifies on what we can coherently believe about the future and our place in the world is often relevant when one is thinking about macrostrategy and similar "big picture" topics. In my personal life, any impact has been more subtle. I think over the years it has contributed to a deepening sense of existential humility and to some dilation of my "spiritual aperture".

10. What should we do if we are in a simulation?

To a first approximation, we should do the same stuff that we should do if we are not in a simulation. I recommend brushing your teeth, getting enough sleep, and being kind to other people and animals.

The original paper has some discussion of various potential practical implications of the simulation hypothesis. Robin Hanson has also published a paper on this. My view on Hanson's paper is that the considerations he points out may well be sound as far as they go, but they are quite weak and in some parts counterbalanced or outweighed by other considerations. The upshot, I think, is that the simulation hypothesis currently does not seem to have any *very radical* implications for how we ought to live, although there might be a multitude of weak or subtle implications. The practical weight of these is further modulated by the possibility that the simulation hypothesis is false.

(Any impulse to try to "hack the simulation" does not seem well-advised. See question 12.)

It is plausible that simulation-related considerations are important in the context of our development of machine superintelligence. They may affect what the superintelligence chooses to do. They also may be both ethically and prudentially relevant to how we should go about creating superintelligence, as I have tried to outline in a recent paper.

References:
Hanson, R. (2001): "How to Live in a Simulation". *Journal of Evolution and Technology*, Vol. 7.
Bostrom, N. (2024): "AI Creation and the Cosmic Host". *Manuscript*.

11. Isn't the simulation hypothesis unfalsifiable?

There are clearly possible observations that would show that we are in a simulation. For example, the simulators could make a "window" pop up in front of you with the text "YOU ARE LIVING IN A COMPUTER SIMULATION. CLICK HERE FOR MORE INFORMATION.". Or they could uplift you into their level of reality.

We could also obtain indirect evidence for the simulation hypothesis. The simulation argument shows that at least one of three propositions (1)-(3) is true. Consequently, evidence that makes the first two propositions less likely would make the third proposition more likely—and conversely. For example, if we discovered that there are hazards on the path towards technological maturity that are so hard to avoid that practically every sufficiently advanced civilization gets destroyed by them, that would increase the probability of (1), and reduce the probability of (2) and (3). By contrast, if we discovered that there doesn't seem to be any such hazard—for instance, because our own human civilization starts to closely approach technology maturity with no clear peril in sight—that would reduce the probability of (1) and hence make (2) and (3) more likely.

So the simulation hypothesis is clearly empirically testable in the sense that there are possible observations we might make that would either increase or decrease the probability that it is true. Further clever analyses might reveal additional ways in which the simulation hypothesis is probabilistically linked to various observable features of the world.

One might wonder whether there is any kind of test that we could conduct in the near term that would definitely reveal which of the three disjuncts is true. I doubt that there is any simple experiment—like mixing some chemicals in a bottle and checking whether they turn red or blue—that would do that. Most theoretical science is of course untestable in *that* sense, so it is not a useful criterion for whether a theory is worth taking seriously.

12. What do you think of doing physics experiments to find out whether we are in a simulation?

Some people have proposed empirical tests of the simulation hypothesis. For example, Beane et al. (2014) propose looking for artifacts that could result if spacetime were discretized on a lattice for the purpose of simulation (anisotropies in the ultra-high-energy cosmic rays distribution). However, there is little reason to suppose that the hypothetical superintelligent simulators would use the crude simulation technique that such a test would detect. As the authors themselves note, modern lattice quantum dynamics simulations run by human physicists routinely use improved lattice techniques that remove this kind of artifacts.

Aside from the specifics, it would seem wildly computationally profligate to base a simulation on a uniformly spaced lattice grid covering the entire observable universe! Presumably, the simulators would instead use techniques that rely more on intelligent processing (generative AI) to create plausible sensory impressions for the simulated beings based on relatively crude representations of their immediate environments.

Another example of a proposed physics test derives from the Ringle & Kovrizhin (2017) attempt to show that classical computers cannot efficiently simulate certain quantum systems (the authors focused on limitations of quantum Monte Carlo simulations in the context of systems with certain topological features linked to the quantum Hall effect). Some media reported this as "scientists have found proof that we are *not* living in a simulation!". The most obvious flaw in that interpretation is that simulators could use quantum computers. More generally, the previous point applies: that while one way to produce the appearance of a given system of physics could be to run a simulation that faithfully computes all the physics in question, another way would be to run a simulation that "cheats" and that produces the appearances in a more intelligently creative or klugey manner: generating patterns that are merely sufficiently realistic in their broad contours to be indiscernible from underlying reality by our primitive human brains (including if necessary by manipulating readings from measurement devices and data files etc.).

If there actually *were* some experimental test we could perform that would reveal whether or not we are in a simulation, it may be wiser not to do it. Preston Greene (2020) has argued that if we discover empirical evidence that we are in a simulation, this would cause the simulation to diverge from basement-level histories. If the rationale for running the simulation depends on the simulation not systematically deviating from non-simulated civilizations (for instance, because the motive is to investigate how non-simulated civilizations develop), then the value of the simulation would decline if its inhabitants discover that they are simulated, and this could lead the simulators to shut it down.

I share the view that if there were some experimental test we could do that would give us clear experimental evidence one way or the other, we should probably not do it. Fortunately, for the

reasons stated above—and especially the consideration that the simulators could fake the results in ways that at our current levels of capability we would be unable to detect—it looks difficult to design such a test. (Likewise, any impulse to try to "hack the simulation" does not seem well-advised. It might be compared to attempting to cause a nuclear explosion by smashing two rocks together—something which seems foolish because it has no chance of working and which would seem even more foolish if it did have a chance of working.)

Note that "philosophical" methods of probing the simulation hypothesis—such as the simulation argument itself—are immune to the concern that Greene points to. This is because both simulated and non-simulated civilizations would engage in such reasoning and would draw the same conclusions from it, unlike potential experimental tests discussed by Greene. Thus it would not have any tendency to cause a divergence between simulated and non-simulated realities.

References:

Beane, S .R., Davoudi, Z., Savage, M. J. (2014): "Constraints on the Universe as a Numerical Simulation". *The European Physics Journal A*, Vol. 50, No. 148, pp. 1-9.
Ringel, Z. & Kovrizhin, D. (2017): "Quantized gravitational responses, the sign problem, and quantum complexity". *Science Advances*, Vol. 3, No. 9, pp. 1-7.
Greene, P. (2020): "The Termination Risks of Simulation Science". *Erkenntnis*, Vol. 85, pp. 489-509.

13. What is the relation between simulation theory and religion?

It does not seem to have any direct logical connection with religious conceptions of a literally omniscient, omnibenevolent, and omnipotent deity. The simulation hypothesis does not imply the existence of such a deity, nor does it imply its non-existence.

The last section of the original paper speculated about certain parallels that could be drawn between traditional religious conceptions and our relations to our hypothetical simulators. These simulators would have created our world; they would be able to monitor everything that happens here; and they would be able to intervene in ways that conflict with the simulated default laws of nature. Moreover, they would presumably be superintelligent (in order to have been able to create such a simulation in the first place). An afterlife in a different simulation or at a different level of reality after death-in-the-simulation would be a real possibility. It is even conceivable that the simulators might mete out rewards to their simulated creatures based on how they behave, perhaps in accordance with familiar moral or religious norms (a possibility that gains a little bit of credibility from the possibility that the simulators might be the descendants of earlier humans who recognized these norms).

One person who had been a hardcore atheist his whole life, told me, when I explained the simulation argument to him, that it was the best argument for God's existence that he had ever heard. He became an agnostic on the spot.

However, it is important to stress that the simulators implied by the simulation hypothesis would be naturalistic entities, subject to the laws of nature at their own level of reality. They would not be strictly omniscient or omnipotent, and they might well be finite. They may all be created by and subordinated under an infinite deity of the supernatural kind envisaged in the great theistic traditions. (Moreover, we need to keep in mind that the simulation argument does not imply the simulation hypothesis.)

Some philosophers have argued that the simulation hypothesis provides new resources to theism by offering a potential response to the problem of natural evil. Since ancient times, people have argued that if theism is true, God would have both the motivation and ability to remove all evils, such as natural disasters, that do not result from the free will of some person. Yet apparently natural evils exist. So, the argument goes, theism cannot be true. One traditional reply is to say that natural evils exist because demons with free will deliberately create them. However, there is not much evidence that demons create all natural evils. Some philosophers have argued that the simulation hypothesis makes a free-will-based theodicy more plausible. If our world is simulated, all apparent natural evils may in fact be the result of choices made by simulators. The base reality created by God may, for aught we know, be devoid of natural evil.

References:
Dainton, B. (2020): "Natural evil: the simulation solution". *Religious Studies*, Vol. 56, No. 2, pp. 209-230.
Crummett, D. (2021): "The real advantages of the simulation solution to the problem of natural evil". *Religious Studies*, Vol. 57, No. 4, pp. 618-633.

14. What if we are simulated by aliens?

"Isn't the argument too anthropocentric? What if some alien species constructed the supercomputer that simulates our universe, and the simulation was made to simulate mainly their kind? In this double-what-if scenario (we are simulated, but by an alien species that has nothing to do with humanity), what are the possible implications to your original theory?"

Formally, the simulation hypothesis includes the possibility that we are simulated by an extraterrestrial civilization. However, the inclusion is redundant. If the simulation hypothesis is true, then we are living inside a computer, and whichever civilization built that computer is our "home" civilization by definition.

Of course, it is possible that the simulators and their ancestors are more similar to some extraterrestrial civilization in our universe (if there are any) than they are to us, so in that sense it is possible that we are simulated by the descendants of an alien-like civilization.

More generally, simulators might create many simulated people who are very different from their own ancestors, or who live in worlds that are very different from the one that their ancestors lived in or the one that the simulators themselves live in. It is possible that we are living in such a simulation. It is unclear how we could estimate the probability that our hypothetical simulators (or their ancestors) are similar to us, or that their world is similar to the world we experience, other than perhaps by speculating on the motives for creating simulations. (The original paper focuses on ancestor-simulations because the methodology is more solid for that case. It is less clear whether some kind of principle of indifference could also be applied to a reference class of "observer-moments" that are very different from one another. For more on the reference class problem, see *Anthropic Bias*.)

References:
Bostrom, N.: *Anthropic Bias: Observation Selection Effects in Science and Philosophy* (Routledge, 2002).

15. What if one is Elon Musk?

Or more generally, "Do some people have more reason to believe the simulation hypothesis than others?". This is not actually a frequently asked question, but perhaps it should be.

Imagine that you are some extremely high-profile or otherwise exceptional or influential individual. In your reflective moments, you might find this fact astonishing. You might think to yourself:

"Out of all the people on this planet, how come that I—me, this little me!—should find myself in such a special position? The odds would seem to be a billion to one against. Should I just accept this as mere coincidence? Or is there some other explanation that would make more sense of my experience?"

Although the epistemology of this kind of indexical reasoning is murky, suppose that you do view the fact of your own extreme specialness as something that would be too improbable to be attributed to happenstance. In that case, a simulation hypothesis might offer you distinctive explanatory resources. It does not seem implausible that simulators would, for one reason or another, be more interested in world-historical or otherwise outstanding figures than they are in your average Joe Bloggs. If not all the simulations they run are simulations of everybody, then the people they find most interesting would be more heavily featured in their partial simulations. For example, such simulations might continuously feature only a small number of key individuals

in sufficient granularity to make them real observers while the extras in such simulations may be simulated only when they interact with the key people or they might be rendered using simplified techniques (such as interpolation of cached statistical patterns) that do not result in conscious experience. If that were so, then a disproportionate fraction of all human observer-moments would belong to high-profile individuals. This could help such individuals make sense of their unique experiences. It would give them an additional (idiosyncratic) reason for thinking that they are living in a computer simulation, on top of the reason that the general version of the simulation argument provides to everybody.

References:
Bostrom, N.: *Deep Utopia: Life and Meaning in a Solved World* (Ideapress, 2024): pp. 165-171. [This section contains some relevant discussion of the technology and metaphysics of NPC characters.]

16. Why do you think there's been so much interest in this?

If the simulation argument is sound, it tells us something surprising and profound about the world. The argument is powerful because from some rather simple and plausible assumptions it derives a remarkable conclusion about the world. It is rare to get so much leverage out of a short argument.

17. If we are merely simulated, does that mean that the world isn't "really real"?

No clarity is gained by asserting that the world isn't "really real" if we are in a simulation. The simulated world that we experience would, however, be only a part of reality. Reality would also contain the computer that runs the simulation, the civilization that built the computer, and perhaps many other simulations and much else besides.

References:
Chalmers, D.: *Reality+: Virtual Worlds and the Problems of Philosophy* (Allen Lane, 2022). [General discussion about metaphysics etc.]

18. How did you come up with this?

In my doctoral research, I studied the foundations of probability theory and self-locating belief, and I worked out the first mathematical theory of observation selection effects (Bostrom 2002). Separately, I had for many years been working on what I've later called macrostrategy, including trying to understand future technological capabilities and their implications and more broadly the big picture situation for human civilization. Combining these two areas, the simulation argument is then only one inferential step away.

Before the idea took its final form, I had for a couple of years been running a rudimentary version of it past colleagues at coffee breaks during conferences. The response would typically be "yeah, that is kind of interesting" and then they would drift on to other topics without anything having been resolved.

One evening—I was a lecturer at Yale at the time—I was again pondering the argument while I was walking to the gym, when it dawned upon me that this was more than some intriguing but amorphous coffee-break material and that it could be recast into a rigorous form. By the time I had finished the physical workout I had also worked out the essential structure of the argument. I got a coffee and went to my office and wrote it up.

(Are there any lessons in this? That new ideas are sometimes born from combining two different areas or cognitive structures, which one has previously mastered at sufficiently a deep level, is well known. I think another moral is that even when we do vaguely realize something, there is often an elusive further step that is necessary for the breakthrough, namely to take the idea seriously enough and not to distractedly let it slip away.)

References:
Bostrom, N.: *Anthropic Bias: Observation Selection Effects in Science and Philosophy* (Routledge, New York, 2002).

19. Can the simulation argument be generalized?

The formal structure of the simulation argument can in principle be applied more generally. For any such application, however, one would need to check whether the empirical prerequisites are in place and whether the result that could be derived is of interest or significance.

For example, we could consider a "terraformation argument", which would be analogous to the simulation argument, except that instead of "living in a computer simulation" it would say "living on a planet that was terraformed and seeded with life by some advanced civilization". The terraformation argument would purport to demonstrate that either almost all civs at our

development stage fail ever to become capable of terraforming, or almost all of those that do become capable decide against, or we almost certainly currently live on a terraformed planet.

If we wanted to make the terraformation argument precisely analogous to the simulation argument, we would need to make several potentially problematic assumptions. For instance, we would need to assume that civilizations could not directly detect whether the planets on which they arose had been terraformed or not. We would also have to assume that each advanced civilization that decided to engage in terraforming would terraform large numbers of planets, and that creatures similar to ourselves would evolve those plants. Furthermore, we would have to consider our temporal position: civilizations arising on terraformed planets would be able, just like we are, to determine the cosmic epoch in which they live, and this might place constraints on how many civilizations arising earlier than ours would have had time to engage in extensive terraforming by the present epoch.

Along similar lines, one could consider a "cosmoformation argument", focusing on the possibility that an advanced civilization using some as-yet unknown technology might be able to induce the creation of baby universes (perhaps by engineering physical singularities that expand into hidden dimensions). The transposition of the simulation argument into a cosmoformation argument (focusing on a "cosmoformation hypothesis", according to which we live in a cosmoformed universe) is straightforward. But the lessons one could learn from such an argument might be somewhat limited. It would not be particularly surprising to learn, for instance, that it is impossible for any ever-so-advanced civilization to spawn new universes. And maybe it would also be less shocking to learn that our universe had been spawned from some parent universe as a consequence of the actions of intelligent agents there than it would be to learn that we are living in a computer simulation.

The bottom line is that terraformation and cosmoformation arguments (or even a "Truman-Show-ification" argument) are compatible with the simulation argument, and they could be cast in a similar logical structure. Some of these alternative applications might be interesting. They each have to be evaluated on their merits.

References:

Dainton, B. (2002): "Innocence Lost: Simulation Scenarios: Prospects and Consequences" *PhilPapers*.