## Introduction

Ancestor simulations are computer simulations of the history of a “simulator” population. There has been much discussion in recent years about whether we, ourselves, are in such a simulation.[[1]](#footnote-2) Preston Greene thinks we should not try to find whether we are in such a simulation. In “The Termination Risks of Simulation Science” (2020) he argues that we shouldn’t run experiments designed to test whether we are in an ancestor simulation because such experiments could result in our termination. His argument runs as follows. A principal motivation for running simulations is to investigate counterfactuals about the history of the simulators. But if the inhabitants of an ancestor simulation discover they are in a simulation, and that discovery significantly impacts the course of events, then because the ancestor simulation no longer suits the simulators’ research purposes, the simulation could be terminated. So, while a successful **simulation probe**—an experiment designed to determine whether the world is a simulation—might contribute to our knowledge of reality, it could also result in our termination. Meanwhile, an unsuccessful simulation probe would not contribute to our knowledge of reality. Thus, given the cost of termination far outweighs the benefit of knowing whether or not we are in an ancestor simulation we should *not* conduct simulation investigations.

In this brief note we argue to the contrary. We do this by way of a series of nested arguments. First, we argue that the simulators are as likely to care about simulating simulations as they are to care about simulating so-called **basement worlds** (worlds which are themselves not simulated, and which contain simulations and possibly nested simulations), and that the simulation discovery would be very relevant to their interests. If we are right about this, then it’s (almost) the end of the matter. But of course, perhaps not everyone will be persuaded, so we move on to consider things on the assumption that the simulators care only about ordinary events that might take place in a basement world. In this section we argue that discovering we are in an ancestor simulation will have little or *no* impact on ordinary events. If ordinary events are not impacted, then we will unlikely fall foul of the simulators’ research purposes, and so discovering we are a simulation will be unlikely to result in our termination.

In the last main section, we consider what to say if we are wrong that the consequences of a simulation probe will make little or no difference to the macroscopic evolution of a simulated world. We argue that the effect of seeking to empirically discover we are a simulation, and at least seeming to do so, are all effects which could happen in a basement world (though of course it would be mere seeming in that case) and thus likely to be of interest to simulators, insofar as they are interested in counterfactuals about the basement world. In the final nested argument in this series of Demosthenic “even ifs”, we consider what to say if all the previous assumptions are, contrary to what we have argued, incorrect. Could there be some kind of catastrophic effect from actually (*veridically*) empirically discovering that we are a simulation which is itself disastrous, or changes the laws of state evolution so as to make the simulation of no interest? We argue this too is unlikely, and that the precautionary principal doesn’t allow one to leverage the low probability.

Every one of our nested arguments would have to be mistaken for it to be true that an ancestor simulation would be put in danger by simulation investigations.

We think our response to this argument is especially important given the cultural traction that these arguments have achieved. The simulation hypothesis has received widespread attention amongst philosophers, physicists, and the general public.[[2]](#footnote-3) Greene’s argument about the termination risk associated with a successful simulation probe is similarly gaining cultural currency and was featured recently in the New York Times (2019).[[3]](#footnote-4)

## 2. Do the simulators only care about what could happen in a basement world?

Let’s begin. Most of the arguments for why simulators would be unhappy with the discovery that we are a simulation depend on different versions of the thought that they would be especially interested in simulating only what could go on in a basement world. The thought is that they care about what could happen in a world like theirs, and their world is a basement world.

However, we think that it is not clear we should just assume the simulators are in a basement world. And if they aren’t, then perhaps they are *more likely* to be interested in simulating what happens when there is a successful simulation discovery, than in simulated societies that never discover this.

Here is why. The hypothesis that we are a simulation is plausible only if simulation is something which is very likely to emerge, and if this is so, there are very likely multiple levels of nested ancestor simulations.[[4]](#footnote-5) Of course, computational limits affect how many levels of simulation can be supported but we have very little idea what these limits are. It will depend on the extent to which the simulators can dispense with computationally expensive detail, and, in a world like ours, how effective quantum computing will be. It will also depend on the laws of nature in the basement world, and for many purposes of simulation it’s unclear that they should be expected to be the same as the apparent micro-laws in simulations. Thus, we should have very little clue how many iterated simulations are possible. But we should think that there are likely as many layers of simulation as is affordable: after all, counterfactuals about what happens to a society when it begins to construct simulations would be a very interesting topic to a society of simulators.

How many layers should we think there are? There are any number of hypotheses to choose from. But only if there are fewer than three layers of simulation, is it likely that if we are in a simulation, then our simulators are in the basement universe. For if there are two layers, and we are in a simulation, then the simulators must be in a basement world. But if there are three layers, then there will a fifty percent chance, and as the number of layers increases (all on the assumption that if we are in a simulated world in a hierarchy, our position is arbitrary), the probability that the simulators are in a basement world decreases.

So, if our simulators are simulations, and either know or strongly suspect they are, then their interest might be very much aligned to finding out the counterfactuals about simulation discoveries. In fact, such simulators might even terminate simulations in which the simulation discovery looks like it’s not going to be made. It is hard to evaluate these probabilities, and this is a thought we will return to at the end of the paper.[[5]](#footnote-6)

## 3. Observation divergence and the importance of successful probes

So far, we have argued that the simulators might be just as interested in simulating a non-basement world as they are in simulating a basement world. But what if Greene is right that the simulators are more likely to be interested in simulating a basement world? In that case the worry is that a successful simulation probe will lead to *observational divergences*; different outcomes from the outcomes which would have occurred had the probe been performed in the basement world. In the next two sections we will explore the issues raised by such divergences, but first we need to introduce some distinctions between different respects in which we might call a simulation probe successful.

Let us say a simulation probe is a **causally successful** simulation probe when:

(a) There is some counterfactual, say *A* → *B*, which would be true in a target basement world, but which is false in a simulation of that world.

(b) *A* is performed in the simulated world and *B* does not occur.

Further, a simulation probe is **apparently successful** when all the macroscopic consequences that C would have occur (different meter readings, and so on).

Finally, a simulation probe is **fully successful** if it is causally successful, apparently successful, and in addition there are probers who know that if we were in basement world, B would occur, and know as the outcome of the probe that it does not.

The remaining sections of the paper consider whether the divergences which could arise because of a fully successful probe pose a termination risk.

## 4. The irrelevance of the simulation discovery for most counterfactuals

Our first pass is that a fully successful simulation probe is unlikely to reduce the overall interest of macroscopic simulations. That is because discovering we are in an ancestor simulation is unlikely to significantly impact upon us and radically alter the course of macroscopic events in our simulation. As Greene acknowledges: “the ancestor-simulation hypothesis is not technically a skeptical hypothesis, but rather a metaphysical hypothesis about the nature of reality.” (pg. 492). We take this to mean that people’s ordinary beliefs about the world would not be made false by the discovery that they inhabit an ancestor simulation, where the fundamental nature of reality is digital as opposed to analog. We would still think that there are trees, clouds, birds, and so on. We would still, plausibly, be *right* about this. We would still want to draw certain distinctions, such as the distinction between free and unfree actions, and would desire that our actions be like those that we judge to be free. We would still be confronted by massive problems that would need to be solved, such as climate change, pandemics, war and so on. The discovery that we are in an ancestor simulation would just be a (very) interesting discovery about the fundamental nature of our world; not a discovery that nothing is real (cf. Chalmers 2010).

If what interests the simulators are ordinary matters, such as Australia in the past going to war with Emus, or whether in the future we solve issues such as climate change, then discovering that we are in an ancestor simulation carries no termination risk. That is because the counterfactual conditions of interest to the simulator are indifferent to being in the basement-level or being in an ancestor simulation. If people’s ordinary beliefs about the world are *not* impacted by the discovery that we are in an ancestor simulation, then people’s ordinary behaviors will *not* be impacted. And, if people’s ordinary behaviors are not impacted by the discovery that we are in an ancestor simulation, then ordinary events will also *not* be impacted. Provided ordinary matters are what interest the simulator, then discovering we are in an ancestor simulation carries no termination risk because it will *not* impact the simulator’s research project at all.

But perhaps we are being a little too hasty. If scientists discover that we live in an ancestor simulation, might some academic interest shift from, say, the study of climate change to the study of the simulation hypothesis, thereby hindering our response to climate change? Similarly, might the discovery we are in an ancestor simulation capture the attention of young people convincing many to become scientists studying simulation theory?[[6]](#footnote-7) We think that the answer in both these cases is still ‘no’. If I am an academic interested in climate science, then to the extent that I am interested in my own climate, why should the discovery that we are in an ancestor simulation cause my interests to shift to simulation theory any more than is already motivated by simulations shedding light on my climate. Similarly, if I am a young person interested in becoming a social worker, then why should the discovery we are in an ancestor simulation cause my interests to shift to science and simulation theory? We see no reason why learning we are in an ancestor simulation should be expected to impact on our interests. At most, you might expect an interest shift within fundamental physics to simulation theory.

Also a fully successful probe would not have a different effect on career choices and so forth from a merely apparently successful probe (one that is apparently successful but not causally successful)*.* This is not a case where there is a difference between successful, empirical discovery that we are in a simulation (which can only take place in a simulated world) and the widespread belief that we are simulations (which can take pace in a basement world). It’s possible for that to happen through faulty science or many other reasons without it being true that we are in an ancestor simulation. This observing what might happen in a fully successful simulation probe in a simulated world, provides evidence for what might happen in the world of the simulators should it become accepted or believed (rightly or wrongly) that they are in a simulation.

One possibility, helpfully stressed by an anonymous referee, is that there might be something special about the effects of a discovery that the world was created by agents: it might undermine our sense of being autonomous, decision making agents. Whether this is so is a matter for empirical psychology, but there are some considerations which make us think it is unlikely.

First, the simulation hypothesis is not one according to which the simulators intervene in the lives of individuals or their decision making. Indeed, were a simulated world not to evolve in the same way as an unsimulated world, it would be useless as a simulation. So, the simulation hypothesis is in this respect, at least, fundamentally like any other deterministic, or stochastic, hypothesis regarding the nature of the world. In fact, one of the authors has some empirical data on what effect various systematic hypotheses about the world being controlled by other agents would have on people’s sense of their own free will, and while we do not put too much weight on this at this early stage, it’s worth noting that the answer appears to be very little![[7]](#footnote-8)

This suggests is that there is no *reason* to lose our sense of ourselves as autonomous agents, able to choose: but of course, sadly, reasons are rarely necessary for psychological effects. Maybe the idea is that if our reality was brought into being by agents, then this might be especially disruptive. But many people, perhaps most, already believe in just such a theory. Theists typically think that the world was brought into being by an agent or agents. How much does this belief affect their behaviour and psychology? Again, this is a very interesting question, but adding more such beliefs to the mix may not be a radical change. It’s very unclear the extent to which, outside theological contexts, the effects of creation beliefs on behaviour are due to the beliefs themselves, rather than the cultural and historical baggage which accompany most religious beliefs. It’s also worth noting, that in the case of some religions, it’s the very thought that our world was brought about by an agent which is in fact crucial to our sense of ourselves as autonomous agents.

Even scientific matters are likely relatively untouched. The special sciences would be unaltered; these describe regularities removed from their implementation in fundamental reality. Even fundamental science might be relatively unaltered: what would happen is merely that its status as fundamental would be in doubt. Physics might remain much the same, it’s just that we would know there is another realm which underlies it. But don’t we already know this, unless we think physics is complete? If a society was on track to discover, for example, faster than light travel, the fact that there is a new but inaccessible layer beneath physics would likely have little effect. Absent any reason to suppose we can have access to that other realm, there is no more reason to imagine people turning from physics to simulation science in droves than any other discovery that there are inaccessible fundamental layers to the universe. Simulation science might be useful to create further layers of simulation. Perhaps also in the bizarre scenario in which you can communicate with your own simulators to ask them to change your own laws of nature, simulation science might be useful too. But if what you are trying to do is achieve faster than light travel in your own world, then the physics that you were developing before you discovered you were in a simulation is what you still need.

## 5. What if the consequences of causally successful simulation probes are more far-reaching?

But now let us suppose that there are relatively large consequences. That is, *C* would not have happened in the basement world, and contra the arguments in the last section, has fairly large effects. Whether this matters depends on the simulators’ interest in certain counterfactuals. The fact that *C* would not have happened in the basement world if *A* were performed is of no consequence if they are interested in what would happen if *C* occurred. Every counterfactual[[8]](#footnote-9) requires a little miracle to happen to bring about its antecedent, so if they are interested in what would happen if *C*, then by the standards of the basement world the miraculous transition from *A* to *C* is one way for it to occur. Of course, if *C* is the kind of thing, which could never occur in the basement world under any circumstances—not just ruled out by determinism and the initial conditions of the basement world—it’s easy to see why it might be of little interest to the simulators. But most of the empirical discoveries which you could make that would make you think that you are in an ancestor simulation are ones which, *at least at the macroscopic level*, would be logically possible in the basement level.

For example, suppose in the basement world there is a lab—Greene Lab A—run by advocates of the simulation hypothesis. Of course, they do not know that they are in a basement world even though they are, and so they aim to test whether their world is a simulation by running a probe. There is another lab in the basement world, Greene Lab B, run by simulation scientists who have high credence that they are in a basement world. They think, therefore, that there can never be a fully successful simulation probe in their world. They are, nevertheless, very interested in counterfactuals about what would happen in a basement world should Greene Lab A come up with convincing evidence that simulationism is true. They might in addition, be interested in what would happen should a simulated world discover that they are in fact a simulation. One way to analyze these counterfactuals is to watch what happens when a lab in the simulation—Greene Lab A(S) (an analog of Greene Lab A in the simulation) — conducts simulation probes. Where Greene Lab A(S) provides convincing evidence they are in a simulation without in fact conducting a causally successful simulation probe (perhaps due to experimental or theoretical error) Greene Lab B gets to test counterfactuals about what would happen if basement world Greene Lab A were to provide such convincing but misleading evidence. Where Greene Lab A(S) conducts a fully successful probe, then they can test what would happen in a simulation should the simulated lab discover that they are in a simulation.

Of course, this second counterfactual is not a counterfactual about the basement world, rather a counterfactual about a simulated world. Insofar as we think the simulators in Greene Lab B are interested in counterfactuals about their world, you might think that they would have little reason to care about this second counterfactual. But giving high credence to the (by our assumption correct) hypothesis that they are in a basement world is not, insofar as they are rational, and understand the possibility of levels of simulation, the same as certainty. Interest in what would happen in a simulated world should it conduct a fully successful simulation probe, seems no less interesting than any other counterfactual. And, if their credence that they are in a basement world is high but not massive, finding out what would happen were they in fact in a simulated world in which a simulation probe was conducted would be interesting indeed.[[9]](#footnote-10) In addition, as we in effect point out in Section 4, this second counterfactual, on the assumption that causally successful probes do not cause massive divergence in evolution compared to merely apparently successful probes, provides information about the first.

We can think of no reason why these counterfactuals should be of lesser interest than others for simulators, and thus no reason why testing them should be more likely to result in termination of the simulated world.

So far, we have largely assumed that the simulators are concerned with simulating things at the macro level of humans and societies. But what if their concerns are not at that level, but instead are at the level of the various physical effects of making a simulation discovery?

At this level, too, apparent discoveries can be things which can happen in a basement universe. Suppose that in a causally and apparently successful simulation probe (to use a laughably simple example) needle *N* comes to point to *S*. In a basement world, a merry apparently successful probe could result in needle *N* also coming to point to *S* by some accident, error, or quantum fluke.[[10]](#footnote-11)

Of course, this is also still macroscopic: what about the widespread, very small-scale phenomena that might occur in a successful probe? There is always the possibility that the simulators are interested in the effects of these phenomena occurring in the basement world, even if they will ultimately occur for different reasons in the simulation. The divergence in the simulation is the little miracle that allows the simulators to test counterfactuals about the basement world. And, these counterfactuals might not be ones about people’s beliefs or wider cultural effects—instead, they might be ones about certain microphysical events occurring in the basement world that in fact were not going to occur in the basement world. One way to test this is to watch them happen in a simulated world, as the simulation probe pulls the course of evolution away from what was deterministically going to happen in the basement world.[[11]](#footnote-12)

## 6. What if simulation probes have non-law governed consequences?

The arguments in the previous sections have assumed that the dynamical laws at the microscopic level will be constant in the simulated world (or at least the rules of evolution at whatever level is of most interest to the simulators). But what if this is not true? What if a causally successful simulation probe, by pushing the simulation off to uncharted territory, causes the universe to turn to gas or ooze or shut-down due to a software glitch?

We don’t have much to say about this case, except that there is no persuasive argument we can think of for thinking that this would occur. It would be poor programming if simulation probes had this effect—and poor programming could result in simulations failing in all sorts of circumstances and there’s no reason for supposing causally successful probes are special in this regard.

The other possibility is that something would change about the laws of evolution in the simulation. We have been assuming that even if the outcome of a probe (*C*) is not what would happen in the basement world, and thus the states of the simulation diverge from the basement world, they diverge by the same dynamical laws or macroscopic laws (in the case of simulations with little micro detail). If, on the other hand, the laws of state evolution change, even without just turning to gas, or ooze, or crashing, then this could compromise the usefulness of the simulation. For the simulation would no longer be a test of what would happen in the basement world if *C*, and instead would just wander off into uncharted, possibly uninteresting, portions of state space.

But again, it seems like there is a burden of proof, or perhaps more leniently, of plausibility, on the part of those who are worried about the simulation. Why should we expect a simulation probe (which of course produces an event which would not have happened in the basement world) to also change the (apparent) laws of nature, or macroscopic laws? Once again this would be an instance of terrible programming. And, if terrible programming is the issue, then we (if we are a simulation) are surely no more at risk from a simulation probe (which is perhaps the kind of very interesting event the programmers have concentrated on and are less likely to have made errors in respect of) than any other circumstances.

One last possibility is that a simulation probe might, in virtue of producing micro changes, result in unpredictable consequences at the macroscopic level. So, suppose the probe changes one subatomic fact, which in turn is what causes the meter reading and the ultimate belief that we are in a simulation, but in turn also has other effects which drastically affect history independently of the macroscopic effects. But for this to be true, there must be tons of true counterfactuals of the form “if this particle had decayed later, history would be drastically altered”. That is a strong assumption indeed, even given the butterfly effect. It would also require the simulators have no interest in those same counterfactuals for it to move the simulation into territory in which they have no interest.

## 7. Conclusion

Perhaps, reader, you remain so far unconvinced and do not think we have ruled out the possibility that the changes induced by a simulation probe will be of no interest to the simulators, or that it could have terrible consequences. We don’t think this is remotely likely, but perhaps it’s possible. Perhaps the world both does have extraordinarily strong butterfly effects, and the simulators have no interest in how the world would evolve given tiny differences in initial conditions. In which case, perhaps, a strong version of the precautionary principle would advise not doing it.

But if we are relying on a strong precautionary principle, then it’s hard to avoid our very first point: that maybe what the simulators really care about is seeing the outcome of a simulation probe for whatever reason, including the likelihood or possibility that they are themselves in a simulation. You don’t need to accept the arguments of Section 2, but merely grant their possibility, for the precautionary principle to recommend in this case avoiding *not* performing the probe, and thus recommend performing it. The principle thus offers no advice, as it recommends both not doing it and doing it. So, the mere possibility of the dangers discussed in Section 6 is no reason to avoid simulation probes.

There is no reason to fear discovering that we are in a simulation, at least insofar as it might lead to our termination. Either it makes no significant difference (as we suspect) to what simulators care about, in which case there is no extra risk of termination, or it does make a difference. And, if it does make a difference, then the difference is likely to be one that the simulators might also be interested in. Finally, the bare possibility of the strange consequences of simulation probes can’t be leveraged via a precautionary principle to advise not probing, because the principle equally advises probing for reasons outlined in Section 2.

So, let’s probe away, should one be developed. It’ll be interesting.

**References:**

Beane, S. R., Davoudi, Z., and Savage, M. J. (2014). Constraints on the universe as a numerical simulation. *The European Physical Journal A* 50(148).

Birch, J. (2013). On the ‘simulation argument’ and selective scepticism. *Erkenntnis* 78: 95-107.

Bostrom, N. (2003). Are we living in a computer simulation? *The Philosophical Quarterly* 53: 243-255.

Bostrom, N. (2009). The simulation argument: Some explanations. *Analysis* 69(3): 459-461.

Chalmers, D. J. (2010). The Matrix as metaphysics. In *The Character of Consciousness* 455-479. Oxford University Press.

Greene, P. (2020). The Termination Risks of Simulation Science. *Erkenntnis* 85, 489-509.

Greene, P. (2019, August 10). Are We Living in a Computer Simulation? Let’s Not Find Out. *The New York Times*. Retrieved from <http://nytimes.com>.

Latham, A. J., and Tierney, H. (2021). The Four-Case Argument and the Existential/Universal Effect. *Erkenntnis*.

Latham, A. J., and Tierney, H. (forthcoming). Defusing Existential and Universal Threats to Compatibilism: a Strawsonian Dilemma for Manipulation Arguments. *Journal of Philosophy*.

Lewis, D. (1979). Counterfactual Dependence and Time’s Arrow. *Nous*, 13(4): 455-476.

Weatherson, B. (2003). Are you a sim? *The Philosophical Quarterly*, 53: 425-31.

1. See for example, Birch (2013), Bostrom (2003, 2009), Chalmers (2010), and Weatherson (2003). [↑](#footnote-ref-2)
2. The seminal paper on the simulation hypothesis is Bostrom (2003). For possible experiments which could be used to test whether or not we are an ancestor simulation see Beane, Davoudi, and Savage (2014). [↑](#footnote-ref-3)
3. Greene also argues that running an ancestor simulation carries a termination risk on broadly computation grounds because of the load on hypothetical computers. In addition, running an ancestor simulation can cause “an unbounded explosion of computational demands if left unchecked” (pg. 503) as those ancestor simulations run their own ancestor simulations, and so on. We won’t address this part of Greene’s paper here. [↑](#footnote-ref-4)
4. Of course, as Greene argues earlier in the paper a basement-level civilization might terminate an ancestor simulation as soon as it attempts to run its own ancestor simulations due to processing limitations. But even if that is right it still seems just as likely that the basement-level simulator would be concerned with us successfully discovering we are in an ancestor simulation versus not discovering we are in an ancestor simulation. [↑](#footnote-ref-5)
5. Preston Greene has suggested that even if the simulators know they are themselves in a simulation, they might nonetheless have interests that would be derailed should their simulation discover that they are in a simulation. This is no doubt true; if they are interested in the results of various ways a mid-21st century pedophilia scandal might go *before the world has discovered itself to be a simulation* (to use Greene’s example) this might get disrupted by the discovery in their simulation that it is a simulation. But there are countless things the simulators might be interested in, and for each interest different developments in the simulation might be a problem. If they were interested in the discovery that the simulated world is a simulation, a pedophilia scandal emerging in their simulated world might disrupt things in a way that masks the usual effects of the simulation discovery! [↑](#footnote-ref-6)
6. We would like to thank Preston Greene for posing these useful questions. [↑](#footnote-ref-7)
7. Recently, [Blinded] found that in cases where one person is being determined people judge that the determined person’s freedom is undermined. However, in cases where everyone is being determined people judged that people’s freedom is *not* undermined. This pattern of responding was found whether people were determined by the past facts and laws of nature, community upbringing, or alien control devices. [↑](#footnote-ref-8)
8. See Lewis (1979). [↑](#footnote-ref-9)
9. Of course, this would be no test of certain counterfactuals: namely whether the simulators would turn off a simulation which conducts a fully successful probe, since the simulators in this case know in advance that they won’t! But it nonetheless might be a good test of other hypotheses about what might go wrong should a successful probe be conducted, such as those we discuss in the next section. [↑](#footnote-ref-10)
10. On the assumption that quantum mechanics is even true of the basement word—if you thought that quantum mechanics is so weird it can’t be the fundamental basis of reality, then we have already conducted causally successful probes and caused divergences from what would happen in the basement world, just very few people have noticed. [↑](#footnote-ref-11)
11. We assume determinism in the basement world here and leave it as an exercise to the reader to make the complications required for an indeterministic basement world. [↑](#footnote-ref-12)