Observer selection effects

Anders Sandberg, Future of humanity Institute, University of Oxford

Selection bias

In science, we want to select samples – individuals, groups, pieces of data – so that the sample obtained is representative of the population we are sampling from. Typically this happens through randomization. But it is possible to select samples in a way that makes it *look* like they are random, yet they actually do not represent the population, giving us a biased picture of what it is like. This is selection bias[1]. It is far more common than it should be (Wainer 1998).

This bias can either be accidental errors (being random is surprisingly hard, too small fishes escape the capture net), or deliberate (cherry-picking data, ending a trial when the conclusion looks good) tilting.

With a little bias from my friends

If I ask my circle of friends what languages they speak, I will get far more responses for English and Swedish than in the true global population. In fact, if you use your friends almost any measure is likely to be biased: we associate with people like ourselves, so they are likely to be like you in many respects.

An important and nontrivial bias among friends going the other direction is the friendship paradox, noted by Scott L. Feld (1991): your friends have on average more friends than you! People with greater numbers of friends have an increased likelihood of being observed among one's own friends.

If you have an undirected graph of friendships (or sexual relations, business interconnections, co-authors (Eom & Jo 2014), twitter followers (Hodas et al 2013), or anything else) with vertices V, edge set E and vertes degree d(v), the average number of friends of a random person in the graph is $\mu=\sum d(v)/|V|=2|E|/|V|$. The average number of friends of friends however is $(\sum d(v)^2)/2|E|= \mu+\sigma^2/ \mu$ since each individual is counted as many times as they have friends. Since typical graphs have high degree variance this can be significantly higher (Grund 2014).

Selecting what to study

There can also be a meta-level problem. Investigation and experiments are done on topics that seem to merit investigation and are amenable to experiments: hence the typical results in a domain will be biased compared to the results that would have happened if we investigated everything in the domain.

---

[1] Sampling bias is sometimes described as a subtype of selection bias, but they are sometimes used synonymously (or just confused). One way of looking at it is to say that sampling bias means the study will have low external validity – the results do not generalize to the rest of the population from the sample - while selection bias means that there is low internal validity – the study does not allow you to distinguish between alternative hypotheses as well as it should. Of course, not knowing about the bias means your conclusions will be wrong and/or overconfident. (Berk 1983)

Fractals shapes are everywhere, but where mathematically analysed and described mainly from the last quarter of the 20st century. Up until that point lines, spheres and the other basic non-fractal shapes were so predominant that from the perspective of science nature was largely treated as an approximation to non-fractals, while fractals were regarded as pathological counterexamples.

Science often looks for keys under the streetlight.

Where to armour aircraft

During WW II aircraft survivability under enemy fire was of obvious importance. Returning planes were scrutinized and bullet holes counted on different parts of the fuselage. How should it be reinforced to improve survivability? Armor could not be placed everywhere, since it would weigh down the plane too much. The statistician Abraham Wald, working at the Statistical Research Group, developed the non-intuitive advice: put the armour where there is no damage on the planes.

Why? Because the planes inspected represented a sample of planes that had managed to return home despite the holes. Planes shot in other locations had not survived, and hence were not represented. Armouring those locations would improve survivability. (Wald 1980, Mangel & Samaniego 1984)

Survivorship bias is a common cause of selection bias. This is one reason so many advertised hedge funds can truthfully show how they have outperformed the market: if the ones that don't are quietly removed, the remaining ones will be successful (and new ones are started all the time) (Ord 2014).

Slow queues

Why does it feel like your queue is always the slower one? If we ignore memory saliency effects (you don't remember short, inconsequential waits) selection effects do play a role: you spend more of your total queue-time in slow queues than in fast ones.

Imagine that there are two queues, and you repeatedly choose one at random. One is twice as slow as the other one. You will hence wait time T or 2T depending on queue. On average, the time spent in the fast queue is $T/(T+2T)=1/3$, and the time in the slow one $2T/(T+2T)=2/3$.

Nick Bostrom did the same argument for traffic queues, pointing out that you should consider "your present observation as a random sample from all the observations made by all the drivers" (Bostrom 2001). Nick also notes that at least in traffic one should switch lanes, since what matters is not the average speed of the lane but rather the speed of the section extending maybe a couple of miles forwards from one's current position. If everybody did this, the speed between the lanes would equalize.

Observation selection effects and anthropic reasoning

These selection biases are example of observation selection effects. An observation selection effect exists when some property of a thing is correlated with the observer existing or being around in the first place.

When such an effect is present the data will be biased, often in nontrivial ways. The more extreme the correlation, the more unusual effects it could have.

For example, all humans – and living multicellular organisms - have an unbroken chain of parents back to the early primordial soup of prokaryotes swapping genes. Yet the probability of any particular lineage surviving across hundreds of millions of years is exceedingly low: 99.9% of all species have gone extinct[2], and your ancestors must have survived at least five major mass extinctions. Yet concluding that we will be equally lucky in the future is erroneous: the past success is no evidence for future success.

Suppose the evolution of life and intelligence requires a set of exceedingly unlikely coincidences: planets at just the right distance from an unusually stable star in the galactic life zone, with a stabilizing moon and a comet-deflecting Jovian, just the right chemical diversity, a fantastically unlikely chemical coincidence producing cells, a long list of low-probability evolutionary steps leading up to a generalist species forced by environmental conditions to become a super-generalist intelligent species. Yet every intelligent species in the universe would have these coincidences under their belt. Conversely, knowing we exist does not tell us whether intelligence is common. (We will return later to the question of whether we can tell whether we are unlikely or common).

Observation selection effects and attempts to handle them is sometimes called anthropic reasoning, from the anthropic principle. The anthropic principle in its simplest form (the Weak Anthropic Principle) states that the laws of physics and conditions must be such that intelligent observers can exist. Stronger, and far more controversial, versions exist (see below). Anthropic bias consists in that estimates of probabilities of events that correlate with the existence of humanity will be biased. Anthropic reasoning seeks to detect, diagnose and cure these biases (Bostrom 2010).

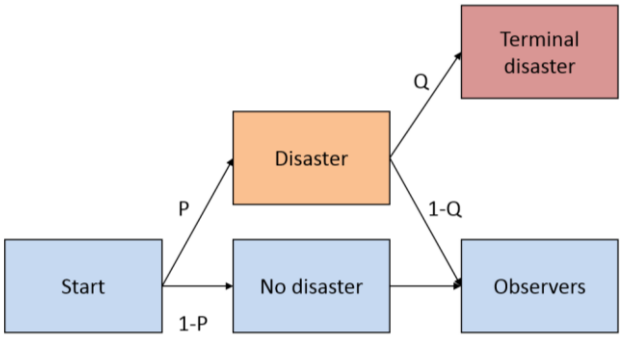
Observer selection effects can sometimes be viewed as looking at causality the wrong way. The existence of observers do not magically change the past or the laws of physics. It only *looks* like that from the perspective of the observers, who of course are in a privileged position. A position they emerged into because of entirely normal causal phenomena.

[2] http://paleobiology.si.edu/geotime/main/foundation_life4.html

Anthropic shadows

If disasters can render planets uninhabitable, we should expect observer selection effects (Ćirković, Sandberg & Bostrom 2010). This also applies if the disasters are not quite deadly.

Why local fossil records are biased



A simple toy model: a disaster either happens with probability P, and if it does it renders the planet uninhabitable with a certain probability Q. Observers only emerge on planets with no disaster (probability (1-P)), or one that was lucky and survived (probability P(1-Q)). If the observers later meet and combine their observations of their joint fossil record they will get an estimated number of disasters far lower than the actual risk. They will see a disaster probability P(1-Q)/(1-P+P(1-Q))=P(1-Q)/(1-PQ), which is off from P by a factor (1-Q)/(1-PQ). The deadlier Q is, the more biased their estimate will be. The world will look safer the more dangerous it is.

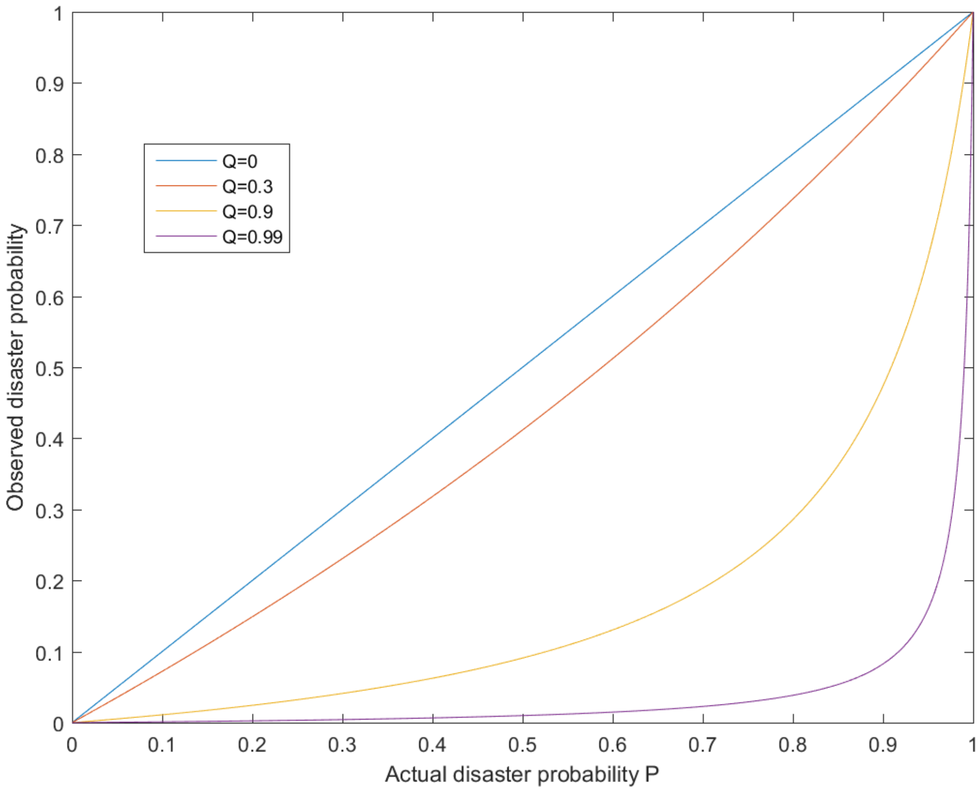


Figure 1: Perceived disaster probability as a function of action disaster probability, for different levels of lethality Q.

This can be generalized to chains of disasters, for example mass extinctions in different eras.

Shadows

Also, if disasters leave temporary but lasting effects on observer appearance there will be a "shadow" of bias.

Consider the ecological effects of a mass extinction: while the event itself can be fast, the recovery of the biosphere appears to take up to tens of mega-years until new ecosystems establish themselves and adaptive radiations allow new species to fill emptied niches (Bowring, Erwin & Isozaki 1999; Sahney & Benton 2008). During this time it is likely that observers are unlikely to appear. Hence observers are less likely to observe a recent mass extinction than they are to observe extinctions in the far past.

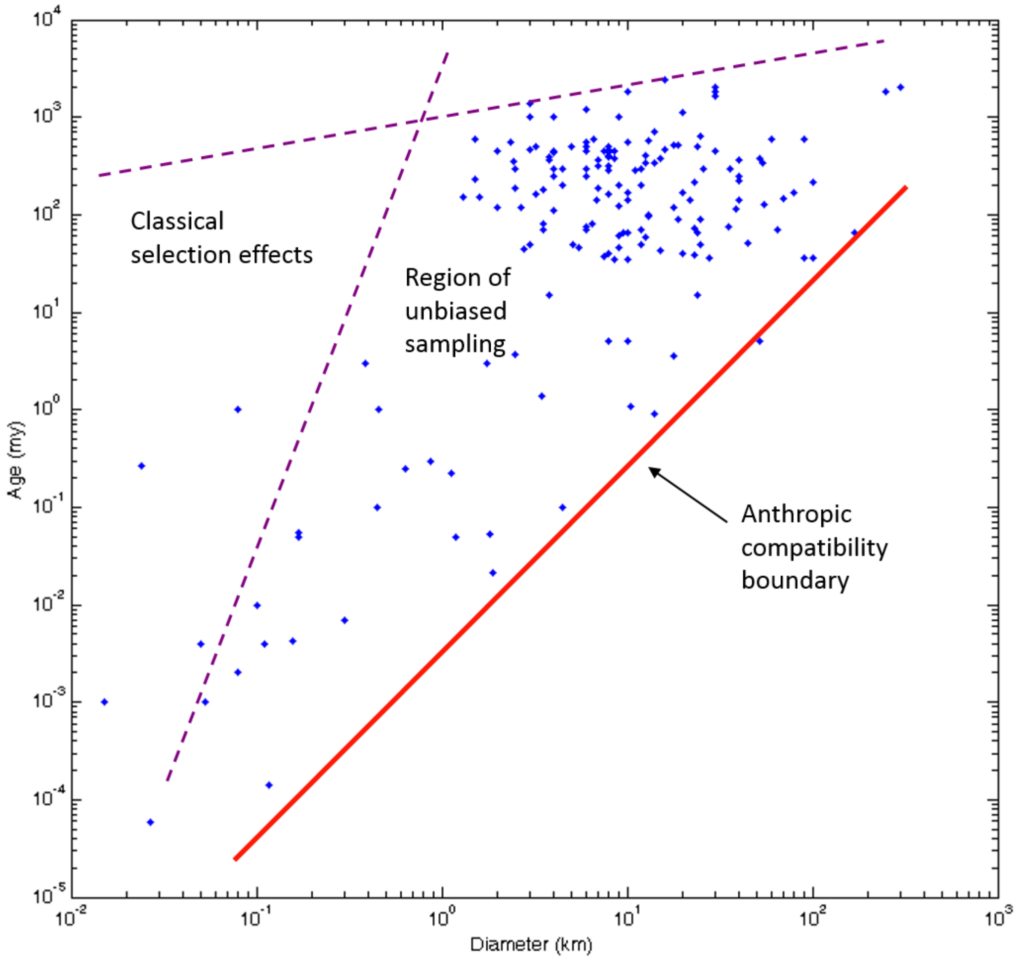


Figure 2: Terrestrial impacts. Horizontal: size/severity, vertical: age. Small and old impacts are underrepresented because of erosion. A speculative anthropic compatibility boundary has been added. Data from http://www.passc.net/EarthImpactDatabase

The existence of such "anthropic shadows" means we cannot rely on the terrestrial crater record to judge the frequency of very severe asteroid impacts, at least not since higher life emerged. However, the moon lacks this bias (beside the bias due to being a different kind of body).

Similarly supervolcano eruptions over the past era may be biased. The largest most recent one, the Toba supereruption between 69,000 and 77,000 years ago, may coincide with a genetic bottleneck implying that our species nearly went extinct (a population down to a few thousand individuals).

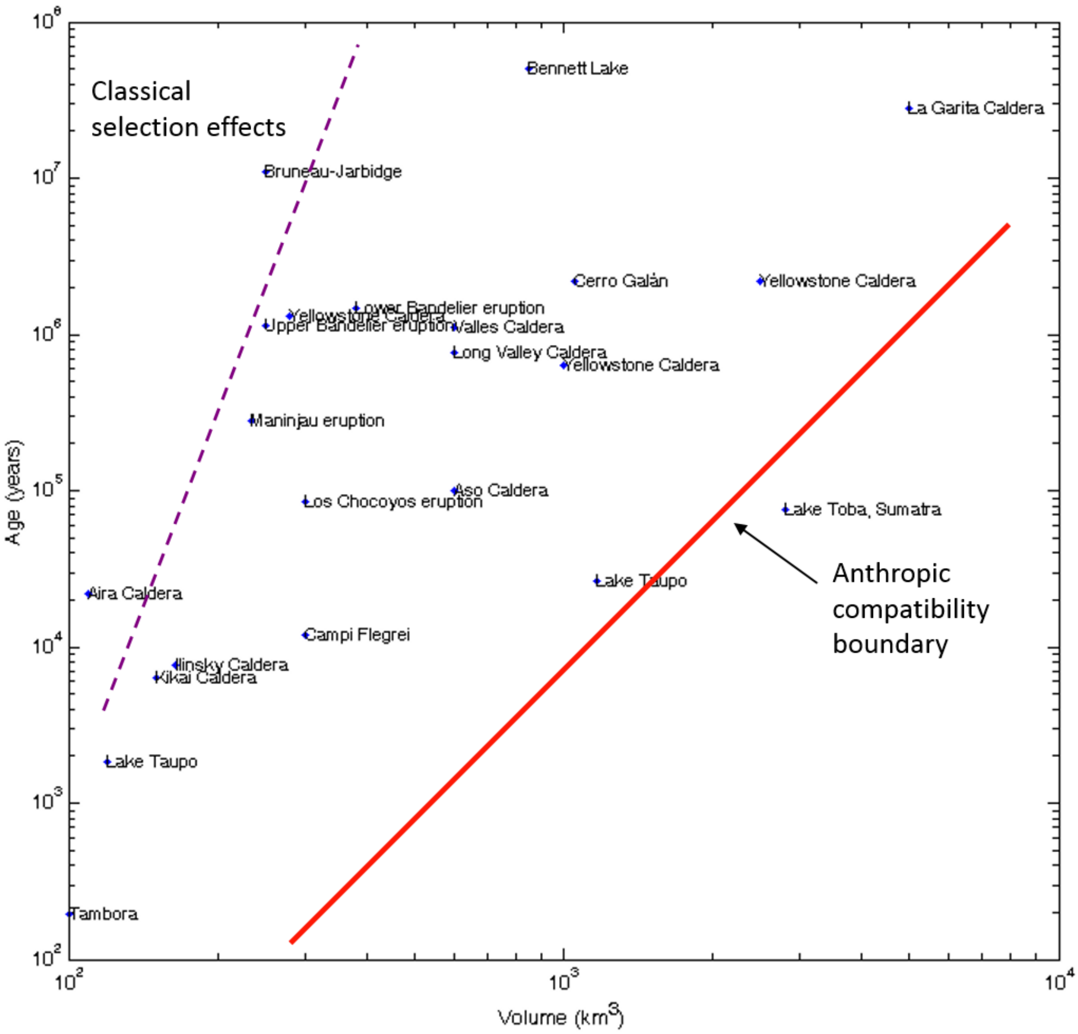


Figure 3: Size and age of supervolism. Source: Wikipedia.

Nuclear war near misses

It is not just natural risks that can be affected by anthropic shadows.

Over the past 70 years there has not been any nuclear war[3]. However, a disturbingly large number of close calls have occurred. These include the 1962 Cuban Missile Crisis[4], the 9 November 1979

---

[3] Absence of war does not mean the probability of war is zero, of course. A simple Bayesian analysis starts by assuming that wars happen independently of each other, that there is some (unknown) probability p of war per year, and as a prior that p is uniformly distributed. Then the probability of n wars after T years is $\binom{T}{n} p^n (1-p)^{T-n}$. Conversely, we can treat this as a (beta) probability distribution $f(p) = p^{\alpha}(1-p)^{\beta} / B(\alpha+1, \beta+1)$. This can be used to estimate likely values of p - even when there are no occurrences of war. If T=70 and n=0 (1945 does not count), then the mean of p is ≈1.4% per year. A 90% confidence interval for p is [0.07%, 4.1%]. This is an estimate not taking observer selection effects into account.

[4] Besides the overall tense situation there was *also* a false alarm at Volk Field, Wisconsin (due to a bear!) that nearly led to nuclear-armed interceptors taking off, an accidental unauthorized US overflight over Soviet

training tape incident (where a training tape was inadvertently inserted into the computer running the US early warning program), the 26 September 1983 incident (where Stanislav Petrov deviated from standard Soviet protocol, avoiding escalation from a false alarm), the 25 January 1995 Norwegian rocket incident (where full alert passed up through the military chain of command all the way to President Boris Yeltsin) – to name just a few (Sagan 1995; Schlosser 2013).

Were a large-scale nuclear war to occur, the effects on humanity would be devastating. While the direct effects would run into hundreds of millions dead, the long-term climate effects ("nuclear winter") would prevent agriculture for nearly a decade globally (Robock, Oman & Stenchikov 2007). While some people close to food stores might survive it would likely massively reduce human populations and increase the risk that other factors bring about species extinction.

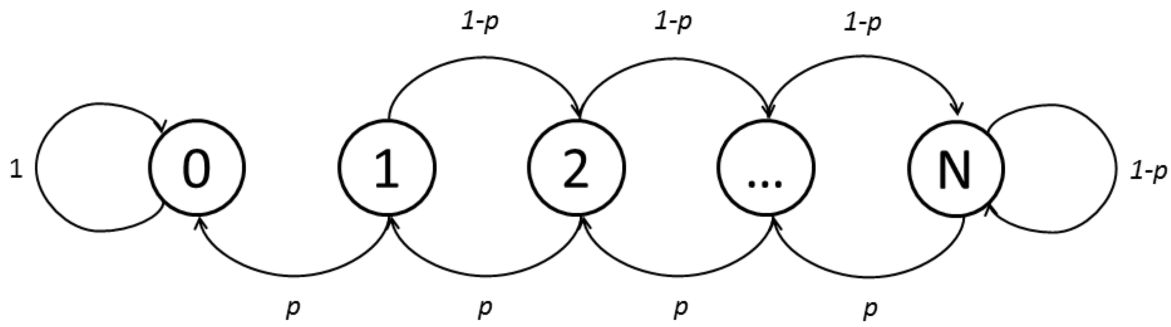


Figure 4: Simple Markov chain model of nuclear war scenarios. The world moves randomly between states of different international tension (N=total peace, 1=serious crisis). State 0 corresponds to an absorbing state with no observers.

If we for simplicity's sake assume nuclear war removes all subsequent observers, then realizations of world history we see cannot have produced a war yet. This can be modelled as a Markov chain (inspired by a model by Martin Hellman (2008)) where the world system randomly wanders between different states of international tension with fixed probabilities of escalation or détente. One state is absorbing (the chain ends there), corresponding to nuclear extinction. Given that at a certain time the system has not been absorbed, what past dynamics does it see?

---

territory nearly led to a confrontation with nuclear air-to-air missiles, a Soviet submarine commander nearly launched nuclear weapons with support from the political officer but was talked down by his second-in-command Vasili Arkhipov, *and* possibly an accidental order to US missile crews on Okinawa to launch.

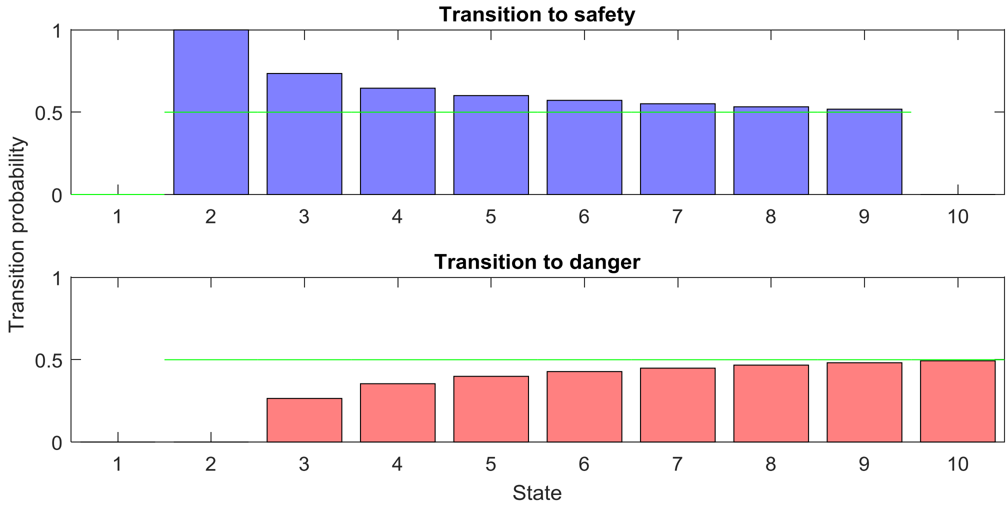


Figure 5: Apparent transition probabilities in the nuclear war model. The true transition probabilities are 50% (green lines), but observers in surviving worlds see biased probabilities implying a far safer world.

It turns out that there is a systematic bias: transitions towards safer states (further away from the risky states that could jump to extinction) are more common than in the underlying unbiased models. Transitions towards greater risk become rarer. This effect is stronger the deadlier the underlying model is: for models with only a minor extinction risk the dynamics is unbiased, while in models where most realizations quickly go extinct observers see highly unusual histories that avoid this by remaining peaceful.

This effect is also stronger in the more distant past: the most recent transitions are less biased. Softening the extinction assumption does not change the qualitative picture much: as long as there are far fewer observers after a nuclear war, history – to the survivors – looks like it avoids it.

The main conclusion is that while the risk estimates for accidental or deliberate nuclear war are unsettling, selection effects may make them underestimates. We cannot be certain of the magnitude of the bias since it depends on the final transition probability to nuclear war, a value that cannot be read out from past realizations. But a priori this argument suggests that we have strong reasons to try to reduce war risk *more* than we would otherwise have tried.

Physics disasters

There have been concerns that particle physics experiments could unleash disastrous effects – stable strangelets, micro-black holes or Q-balls absorbing the earth, or events triggering the collapse of the vacuum state of the universe. Physicists have tried to calm the concerns with both theoretical arguments[5] and the more practical argument that cosmic ray collisions of far higher energy have been going on for eons (Ellis et al. 2008).

Cosmic ray implosions

The cosmic ray argument in its basic form has an anthropic flaw: if Earth had been destroyed by a cosmic ray-triggered implosion we would not exist. So even if this was very likely we might be rare lucky observers and our observations are biased.

A simple save might be to argue that such observers should expect to see a solar system where most bodies but their own planet have been imploded: since this is not what we see, we should reject the danger hypothesis. Unfortunately the energy released by a planetary body imploding is likely enough to sterilize the solar system, so this save will not work.

By looking at supernova rates we may get anthropically unbiased information. Were we to live in a risky universe we should expect to see many distant supernovas (and perhaps an unusually low rate of nearby ones, if they could wipe us out). This way physicists have argued that an unacceptable risk would imply a far higher supernova rate than we see (Dar, Rújula & Heinz 1999; Giddings & Mangano 2008).

Vacuum decay

Even if implosion to compact objects can be ruled out this way, we have another problem: vacuum decay is impossible to see, since it spreads at the speed of light. Astronomical information does not seem to give any evidence to help us estimate the risk.

However, an ingenious argument by Nick Bostrom and Max Tegmark bounds the risk in a sideways way (Tegmark & Bostrom 2005): A world where vacuum decay is easy to trigger will have many independent regions expanding at the speed of light, gradually merging until the entire universe is converted. Observers can only be found in the era before everything is converted. Hence observers will be early in the history of the universe: the probability of being an observer at time t declines as $e^{-t/\tau}$ where τ is the probability per year of annihilation happening.

Assuming observers need to show up on planets, comparing the age of the Earth to what we know of planet formation rates gives us some evidence of whether we are early or late observers. The unbiased probability distribution of formation dates $f(t)$ is turned into a biased distribution $e^{-1/\tau}f(t)$ if we only observe undestroyed planets. The result is a planetary formation curve that peaks earlier and earlier the higher τ is. Given Earth's formation date, we can see that this datum is exceedingly unlikely if τ is shorter than 1 in a billion per year.

---

[5] The theoretical arguments – and many of the practical ones too - have an interesting epistemic problem: they are complex arguments in an uncertain domain trying to bound a low probability, with a higher risk of the argument being flawed than the probability they are trying to bound. They are hence weaker than they at first look, and must be handled carefully (Ord, Hillerbrand & Sandberg 2010).

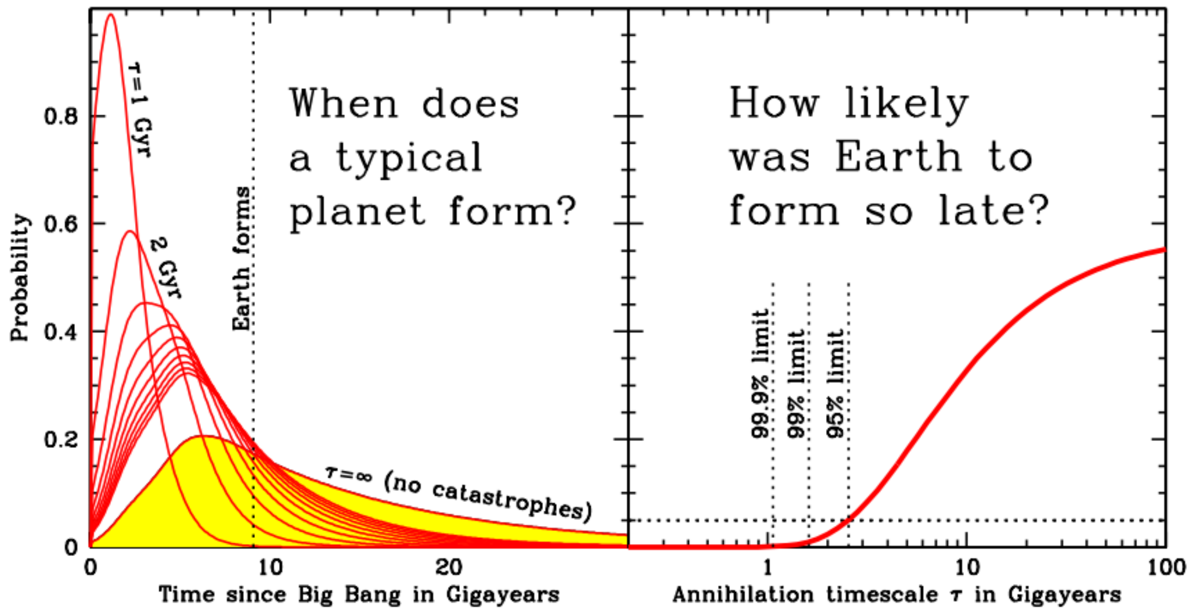


Figure 6: (left) Probability distribution for planet formation given different catastrophe timescales. (Right): probability of observing Earth forming after 9.1 Gyr after the big bang.

The best part about this argument is that it is agnostic about the details of the threat: here we considered vacuum decay, but *any* other threat that occurs randomly and precludes observers afterwards is covered by the argument. We hence have a bound for not just the original risk, but many unknown or unconceived risks!

Is a long-running LHC evidence of safety?

The LHC has been in use since 2008, successfully finding the Higgs boson and doing much useful science. Given that we are still around, the people who worried about the risk were wrong, right?

By now the reader will guess that observer effects will throw a spanner in the common sense works. If the LHC had – through any mechanism – destroyed the world we would not be around to observe it. So even if it was immensely risky we might be very lucky (and wrong about safety).

However, we might again argue as above: if we say the expected usable lifespan of the LHC is 20 years, and we are 35% into it. If it was very dangerous we should expect ourselves to be early observers.

Another argument for LHC safety is the lack of unusual faults preventing its operation. Hans Moravec sketched a scenario where future accelerator builders found their machine impossible to turn on: every time bad luck hit, in the form of independent events like power outages, blown fuses, janitors tripping over cables, etc. Eventually the builders realized that in those universes where the accelerator was turned on vacuum collapse did occur, so only in the remainder would observers remain (Moravec 1988, p. 188).

Indeed, there were delays of the LHC because a quenching incident as well as a power-cut due to a bird eating a sandwich on a capacitor[6] (!). But if the LHC was very likely to cause risk, we should

[6] http://www.theguardian.com/science/2009/nov/06/cern-big-bang-goes-phut

expect these problems to happen far more often[7]. Given that negative funding decisions can be viewed as correlated fairly high-probability faults, the fact that the LHC was funded in fact speaks against it being risky.

[7] If P is the probability that the accelerator is used safely (the universe is destroyed with probability 1-P) and Q is the probability that a fault prevents operation, then a faction P+(1-P)Q worlds survive an attempt to turning it on. Of these worlds a fraction (1-P)Q/(P+(1-P)Q) are saved by the fault. The lower P is the more likely faults are to happen.

The Doomsday Argument

The German tank problem

A classic sampling problem is the “German tank problem”: during WW II the allied forces examined serial numbers on destroyed or captured tanks. The gearbox numbers were consecutive, allowing a statistical estimate of the number of tanks manufactured so far (of obvious military planning importance (Ruggles & Brodie 1947)). Given a tank numbered 217, how many tanks should we expect to have been made so far? What if the number was 1217 or 17?

If you have seen just one tank with number N the answer is about 2N-1 tanks in total[8]. The simple answer is to assume the sample is somewhere in the middle of the set.

The Doomsday Argument

Now consider all of humans, ordered by birth order. Where in this set should you expect to find yourself? It is most likely that you are somewhere in the middle, rather than an early or late outlier. This can be seen as a version of the Copernican principle: we should not expect to be in any particular spot.

There have been about 100 billion people. If we assume the population levels out at 9 billion people and survives for X years with 100 year lifespans, then there will be 9e9*X/100=9e7*X people in the future. If we are exactly in the middle, we should hence expect X=100e9/9e7=1111 years!

Assuming exponential growth makes things even worse. Calculating using 90% confidence intervals do not change things much[9]. Doom within a few millennia, at most[10].

We can use this to consider our chances to become a huge supercivilization. Either we go extinct in a few thousand years, or we spread to the stars, getting billions of billions of descendants[11]. Ought we to hence expect to find ourselves among the first 100 billion out of a billion billion? This seems unlikely (1 in 10 million). So if this reasoning holds, we should not expect a grand future (Gott 1993).

This is the Doomsday argument, independently discovered by Brandon Carter, H. B. Nielsen and Richard Gott, and developed at length by John Leslie (Leslie 1990; Leslie 1998; Bostrom 1998; Bostrom 2010). It can be formulated as a time estimate for a predicted future Doomsday, but the

---

[8] If you are a frequentist. The Bayesian treatment diverges for one tank if one uses a non-informative prior, see (Höhle & Held 2006).

[9] If there will be X people in total, we are with 90% probability in the interval [0.05*X,0.95*X]. If we assume the 100 billion earlier people were not in the interval but we are right at the lower edge, we get X=100e9/0.05=2 trillion. But this is the upper bound. If we assume we are uniformly distributed in the interval the mean of X becomes $<X>=\frac{1}{0.95-0.05}\int_{0.05}^{0.95}\frac{10^{11}}{t}dt=1.1\cdot 10^{11}(\ln 0.95-\ln 0.05)=$ 327 billion. By this calculation we are 30% of the way through history.

[10] This kind of reasoning can be applied to other things, of course. For example, where in the history of Twitter should we expect to be? https://what-if.xkcd.com/65/ If encountering something at a random point in its lifetime, you have a 50% confidence interval that it is going to last between [1/3,3] times its current age, and a 95% confidence interval that it is going to last [1/39,39] times its current age (Gott 1993).

[11] See Nick’s astronomical waste paper (Bostrom 2003). My own estimates are around 14 trillion future lives if we merely survive for 800,000 years as sustainable farmers, $1.75\cdot 10^{16}$ for a billion years, $1.53\cdot 10^{29}$ lives if we colonize Earthlike planets in the milky way, $6.12\cdot 10^{38}$ if we go to other galaxies, $1.02\cdot 10^{47}$ if we live in space colonies, and up to $3.92\cdot 10^{100}$ lives if they are virtual and makes optimal use of the energy in the reachable universe. As noted in the paper, the exact numbers do not matter much, just their overwhelming magnitude.

more philosophically interesting form is where it seems to make a prediction about the size of the future human race, ruling out very large futures on the same grounds as seeing tank 127 makes it unlikely that there are millions of tanks.

Most people respond to the Doomsday argument with "that can't be right!" But *why* (or even if) it is problematic is a large topic with no consensus.

Reference classes

Could the problem be that we are using the wrong reference class? Maybe future posthumans are so unlike us that they do not count as observers, just as we do not count the animal-like pre-humans in the above 100 billion estimate? The Doomsday argument just happily predicts that in short order our species will transcend. But if what matters is being an *observer* rather than being Homo sapiens, why should we find ourselves among the early observers?

What about applying it in the past? Early humans could have applied it to predict that the history of Homo sapiens would be very short. However, if every human makes the argument that they are within the mid 90% interval of human history, only 10% will be wrong. The argument is probabilistic rather than something true for every individual.

Also, as noted by J.R. Gott, one needs to be careful to apply this kind of reasoning to a *random* point in time of the thing studied: when attending a wedding the prediction of how long it will last based on its current length will be wrong, since it is a special time for the phenomenon (Gott 1993). If we have extra information about typical species lifetimes or suspect we are early this could act as a prior changing the conclusion of the argument strongly.

Sampling assumptions

Much hinges on unstated assumptions about how observers are selected.

There is a strong similarity with the Sleeping Beauty problem:

> "Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you back to sleep with a drug that makes you forget that waking. When you are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads?" (Elga 2000)

Some readers will conclude that it is obvious that the probability is 1/2: it is a fair coin after all. Others will equally readily answer 1/3: a third of awakenings will be Heads.

These are mirrored by the two main contenders for how to treat the selection in the Doomsday argument:

> **The Self -Sampling Assumption:** "All other things equal, an observer should reason as if they are randomly selected from the set of all actually existent observers (past, present and future) in their reference class."

Believers in the SSA answer 1/2 in Sleeping Beauty (since there the number of observers is the same in both the Heads and Tails case).

This assumption requires careful handling of the reference class.

> **The Self- Indication Assumption:** "All other things equal, an observer should reason as if they are randomly selected from the set of all possible observers."

Believers in the SIA answer 1/3 in Sleeping Beauty (since there are more observer moments in the Tails case).

Here "randomly selected" is weighted by the probability of the observers existing: under SIA you are still unlikely to be an unlikely observer, unless there are a lot of them. It also predicts a lot of observers: the more observers a hypothesis implies, the more it is favoured.

The SIA avoids the doomsday argument, which is why Ken Olum proposed it (Olum 2002). Whether it actually succeeds is another matter, since it runs into trouble with the Presumptuous Philosopher paradox (Bostrom & Ćirković 2003c). On the other hand, the SSA has plenty of problematic consequences too (Bostrom 2001b).

The Fermi Paradox and the Great Filter

The Fermi Paradox

The Fermi paradox (strictly speaking, the Fermi question – "Where are they?") deals with the apparent contradiction between a large, old universe where numerous species of intelligent life *could* have evolved and the lack of evidence for such beings.

There are about $2.5 \cdot 10^{11}$ stars in the Milky Way and about $5 \cdot 10^{22}$ stars in the visible universe. Planetary systems appear to be relatively common. Assuming the mediocrity principle (Earth is not special, but merely a typical planet, subject to the same laws, effects, and likely outcomes as any other world[12]), it would appear that life and in particular intelligent life would have developed on some of these worlds. Even if a very small fraction such worlds developed intelligence, e.g. 10^{-9}, it would imply hundreds of intelligent species in the Milky Way. They would also be spread out in time: terrestrial planets could have formed billions of years before the formation of Earth, so some of these species would have had many millions, possibly billions of years, to develop and spread across the galaxy. Given that intelligence implies problem-solving abilities and a tendency to overcome scarcity, as well as life's general tendency to colonize new habitats, it would appear likely that at least some of these would have done so. Even at a relatively slow spread would colonize the galactic disk within 50 million to one billion years (Newman & Sagan 1981), a very short time compared to the galactic age of 10^{10} years[13]. Advanced civilizations would both likely be noticeable and lasting. However, we do not see any evidence of past or present visitations in the solar system, nor any signs of technological activities elsewhere. A large number of explanations for the Fermi paradox have been proposed (Brin 1983; Webb 2002).

The conceptually simplest answer is that intelligence is extremely rare, so we are alone in the observable universe. Another simple answer is that intelligence does not tend to last long enough to put its mark on the universe. Knowing the relative probability of the two answers might give important actionable information (e.g. whether to embark on a drastic extinction-avoidance program[14]).

The Great Filter

Robin Hanson (1998) pointed out that the process of detectable intelligence emerging has to pass through a set of steps of varying probability. He gives as an example:

1. The right star system (including organics)
2. Reproductive something (e.g. RNA)
3. Simple (prokaryotic) single-cell life
4. Complex (archaeatic & eukaryotic) single-cell life
5. Sexual reproduction
6. Multi-cell life

[12] By now the reader may be sceptical about this principle. Because of selection effects we might indeed be in an unusual position, e.g. as argued by (Ward & Brownlee 2000).

[13] Allowing intergalactic colonization amplifies the problem by a factor of millions or billions (Armstrong & Sandberg 2013).

[14] If intelligence tends to be doomed, the problem may be that this also applies to *forewarned* intelligence: other species would also have considered the same problem, to no avail.

7. Tool-using animals with big brains
8. Where we are now
9. Colonization explosion

Given the absence of events passing through all nine steps despite hundreds of billions of “trials”, at least one of the steps need to be highly improbable.

If the improbable step is in 1-7 we should expect to have a bright future: possibly alone in the universe, but with a good chance to colonize it and fill it with life and intelligence of terrestrial origin. If it is step 9 that is improbable, either our technological optimism is misplaced, or there is danger ahead.

Were we to discover evidence for (independently evolved) extra-terrestrial life, this would tell us that some of the steps are not unlikely. This would move the credence that step 9 is improbable upwards, i.e. bad news for us (Bostrom 2008). Applying different self-sampling or –indication assumptions to the filter generally suggests extinction risks are underrated (Grace 2010).

Applying Bayesian probability to the Filter shows that evidence (alien life, extinct human-level civilizations) move the credence in a hard step 9 very differently depending on the choice of prior probabilities: step probabilities uniform on the interval [0,1], uniform on a logarithmic scale, or correlated with each other, produce very different responses (Verendel & Häggström 2015). One should hence be cautious about inferring too much from single pieces of evidence.

Does the timing of evolutionary events on Earth tell us anything? Life may have appeared 4.1 billion years ago, and was certainly around 3.5 billion years ago (Bell et al. 2015). Given that the earth formed 4.54 billion years ago, this means that life appeared in the first 9-22% of Earth’s history. This can be interpreted as life is an easy, highly likely step. However, Hanson points out that this could be subject of strong anthropic bias. For a step of probability P, the average time for it to occur scales as 1/P. A very unlikely step might hence be expected to take longer time than the planet remains habitable. But conditioning on the step actually occurring during the habitable period, it will be uniformly distributed in the period. If two unlikely steps have to happen in sequence, then they will both take about the same time, in expectation a third of the period. Easy steps will still occur in time 1/P. Hence Hanson concludes that the early appearance of life might not be evidence that it is likely, merely that there are several hard steps leading up to us.

Other issues

This section deals with more speculative applications of observer selection theory.

The Simulation Argument

Nick Bostrom's simulation argument runs as follows (Bostrom 2003): it seems feasible that advanced civilizations could run simulations of minds and worlds, especially ones like our present ones. Indeed, there are many reasons for future posthuman civilizations to do so (entertainment, resurrecting ancestors, understanding history…) and the resources of an advanced civilization would be enough for a very large number of ancestor simulations. In this case set of minds like ours is dominated by a vast number of simulated minds. Doing the maths properly one concludes that one of the following has to be 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.

The argument applies to similar cases where advanced agents could generate vast numbers of minds[15].

The Anthropic Principle(s)

There are many statements of "The" anthropic principle, many which are contradictory or deal with different topics (Barrow & Tipler 1986). One general definition is: "the drawing of scientific inferences from a consideration of Man's Place in Nature." (Tipler 1988)

Brandon Carter originally coined the term "anthropic principle" in 1974, in analogy with the Copernican principle (that we should not expect to occupy a privileged central position in the universe). He noted that some of the "coincidences" in physics had implications for the possibility of observers (Carter 1974):

> *Weak anthropic principle (WAP) (Carter)*: "We must be prepared to take into account the fact that our location in the universe is necessarily privileged to the extent of being compatible with our existence as observers."
>
> *Strong anthropic principle (SAP) (Carter)*: "The Universe (and hence the fundamental parameters on which it depends) must be such as to admit the creation of observers within it at some stage."

The recognition of principles like this is older; see for example (Wallace 1903, p. 306) and (Dicke 1957): "The age of the universe, "now, " is not random but is conditioned by biological factors".

Barrow and Tipler defined variants of the principles, with their weak principle stating the necessity of carbon-based life rather than just observers and including the dependency on fundamental parameters[16]:

---

[15] http://www.simulation-argument.com/

Weak anthropic principle (Barrow & Tipler): "The observed values of all physical and cosmological quantities are not equally probable but they take on values restricted by the requirement that there exist sites where carbon-based life can evolve and by the requirements that the universe be old enough for it to have already done so." (Barrow & Tipler 1986, p. 16)

Their strong principle includes a "must" with fairly heavy teleological or metaphysical implications:

Strong anthropic principle (Barrow & Tipler): "The Universe must have those properties which allow life to develop within it at some stage in its history." (Barrow & Tipler 1986, p. 22)

Even stronger versions have been proposed, such as Wheeler's

Participatory anthropic principle (PAP): "Observers are necessary to bring the universe into being."

Barrow and Tipler also formulated the

Final anthropic principle (FAP): "Intelligent information processing must come into existence in the universe, and, once it comes into existence, it will never die out."

The WAP is sometimes claimed to be a tautology, but Barrow argues that it does have nontrivial results, such as ruling out certain categories of cosmological models and showing how apparently independent facts about nature (such as the size of the universe and the timescale for biological evolution (Dicke 1961)) have necessary links (Barrow 1983). The WAP is relatively uncontroversial as a statement, although different thinkers disagree on how much content it has or the validity of arguments based on it.

The SAP is far more controversial. The "must" seems to imply some form of purpose of the universe, and can be interpreted as (1) there exists one possible universe "designed" with the goal of having observers, (2) observers are necessary to bring the universe into being (as per the PAP), or (3) there exists an ensemble of other different universes necessary for the existence of our universe (Stenger 2007).

While theists may like (1) there is little content in the principle supporting any particular deity[17]. (2) is not taken very seriously by physicists. (3) on the other hand is popular (if controversial) among physicists and philosophers. It also fits well with the large number of observers favoured by the SIA.

The FAP and PAP are clearly even more radically speculative, and have contributed to the scepticism of mainstream physicists to anthropic reasoning. Tipler built an elaborate theory on the FAP (Tipler 1988), which unfortunately ran into problems with recent cosmological observations.

---

[16] Barrow has an earlier definition of the principle that only talks about variable quantities like age, size and inhomogeneity level of the universe rather than the fundamental quantities like the fine structure constant (Barrow 1983); in this paper he clearly states it does not imply any ensemble of universes.

[17] Indeed, one could imagine the simulation argument writ large as an explanation; see (Steinhart 2010) for a theological example.

Gott defined the Copernican anthropic principle, which says that you should consider yourself as being randomly sampled from the set of all intelligent observers:

> *Copernican anthropic principle* (*CAP*): "[T]he location of your birth in space and time in the Universe is privileged (or special) only to the extent implied by the fact that you are an intelligent observer, that your location among intelligent observers is not special but rather picked at random from the set of all intelligent observers (past, present and future any one of whom you could have been." (Gott 1993, p. 316)

This can be viewed as a restatement of the SSA. Like Carter and Dicke, this makes weak claims about physics and is mostly about the character of observer selection bias. The stronger forms of Barrow, Tipler and Wheeler make claims about physics, metaphysics or teleology.

Fine tuning

The classic use of anthropic reasoning is to argue about the fine-tuning of the universe.

It has long been argued that life would not be possible if the parameters of physics were even mildly different (Davies 1982; Barrow & Tipler 1986; Rees 2000; Stenger 2011; Barnes 2012).

A list of some of the important parameters that have been claimed to be fine-tuned:

Parameter	**Effects**
$(\alpha_e, \alpha_s)=(1/137,0.1)$ (electromagnetism vs. strong force).	Avoid stable diprotons (occurs with a 3.7% increase in α_s (Davies 1972)), unstable carbon (happens if $\alpha_s<0.3\alpha^{1/2}$), deuterium unbound (happens if α_s reduced by 11% (Davies 1982)), and relativistic atoms ($\alpha_e<<1$). (Tegmark 1998)
$(\alpha_e, \beta)=(1/137,1/1836)$ (electromagnetism vs. electron/proton mass).	Avoid lack of stable ordered structures (occurs if $\beta>1/81$), enable star formation ($\beta<\alpha_e^2/200$). (Tegmark 1998)
$m_n/m_p>1+\beta$	Preventing stable neutrons from turning nearly all hydrogen to helium during nucleosynthesis. Had $m_n/m_p<1-\beta$ protons would have been unstable. (Tegmark 1998)
$\varepsilon=(m_{He}-(2m_p+m_n))/m_{He}=0.007$ (nucleon binding force in helium).	If $\varepsilon=0.006$ only hydrogen could exist, $\varepsilon=0.008$: all hydrogen would have fused during the Big Bang (Rees 2000).
$\alpha_e > \alpha_g$, $\alpha_w >> \alpha_g$	Electromagnetism must be stronger than gravity to allow non-collapsed matter, the weak force must be stronger to allow a hierarchy of scales. (Tegmark 1998).
α_w (weak force)	If weaker, all hydrogen converted to helium shortly after the Big Bang, if much larger or smaller neutrinos fail to blow away outer parts of stars in supernova explosions, preventing heavier elements from leave the stars (Tegmark 1998)
$m_e<m_n-m_p$	Electron mass less than the mass difference between protons and neutrons: unstable neutrons do not dominate baryonic mass.
$\Lambda=10^{-122}$ (cosmological constant).	If too large, astronomical structure could not form; the actual

	value is a few times lower than the boundary (Weinberg 1987). The value is also unusually small for a physical constant.
Quark masses in the right ratio.	(Jaffe, Jenkins & Kimchi 2009) find a band about 29 MeV wide in terms of up/down-quark mass-difference (corresponding to ≈10% of the range in the model) that allows the formation of stable atoms.
ν (Higgs vacuum expectation value)	Were ν 1% less, orders of magnitude less carbon would be produced, were it 50% less protons would decay into neutrons. Were it 1% larger orders of magnitude less oxygen would be produced, 2 times larger deuterium would be unstable, 5 times larger neutrons would decay into protons inside nuclei leaving only hydrogen stable. <80% would allow diprotons and dineutrons, upsetting nuclear synthesis. (Tegmark et al. 2006)
$\frac{\nu}{\Lambda_{qcd}} \approx 1$	Ratio of electroweak vacuum expectation value to the QCD scale needs to be about one, or matter-antimatter asymmetry would not occur (Hall & Nomura 2008).
Ω≈1 (density parameter: relation gravity and expansion energy of universe)	If too large, universe collapses early, too small no stars.
$Q=10^{-5}$ (primordial density fluctuations).	Smaller than 10^{-6}: no stars, substantially larger than 10^{-5}: supermassive galaxies with frequent planetary orbit disruption, and above that a too violent universe dominated by black holes. (Tegmark & Rees; Rees 2000). Together with dark energy, photon and matter density it also constrains halo formation (Tegmark et al. 2006).
Initial entropy level	Too low, no stars formed in protogalaxies; too much, no proto-galaxies (Linewaver & Egan 2012)
Spacetime being 3+1 dimensional	Matter stability, information processing, and life appear problematic in higher dimensions or with more time dimensions (Tegmark 1997; Tegmark 1998).
Complex properties	
Carbon production in stars	Carbon 12 abundantly produced because 7.656 MeV Hoyle resonance coincidence[18] with $^{8}Be+^{4}He$ energy (plus ^{8}Be decay lifetime 10,000 times longer than 2 alpha scattering, and conversion of $^{12}C+^{4}He \rightarrow ^{16}O$ harder than conversion to C). (Livio et al. 1989). This is stable for changes in fine structure constant and light quark mass by 2% (Epelbaum et al. 2013)
Water properties	The properties of water (solvent properties, buffering, expansion when freezing, wide liquid interval, high boiling

[18] This resonance is sometimes claimed to be a great success for anthropic reasoning. In 1953 Hoyle predicted the existence of the excited state of carbon since without it little would be formed, and the resonance was soon found. However, the prediction was not based on observer selection considerations at the time: only afterwards was it ascribed to anthropic reasoning (Kragh 2010).

	point, high heat capacity, high surface tension, etc.) are uncommon[19] among chemicals yet essential for terrestrial life. These largely follow from the formation of hydrogen bonds, the 104.5 degree bond angle and a strong dipole moment. In turn, these are nontrivial consequences of the fine structure constant (Lynden-Bell et al. 2010, pp. 123-124).
Earth climate stable	Earth has unexpectedly slow Milankovitch climate cycles (Waltham 2011).

Needless to say, theists love to claim this provides scientific evidence of a creator. Fine-tuning can also be explained given observer selection effects; one should perhaps conditionalize the probability of a parameter value on the number of observers they allow (Weinberg 1987).

But there are at least three alternatives or complicating factors:

First, we might simply not have enough imagination to consider how life could function in a significantly different physics[20], so we are being parochial about how essential these parameters are. For example, by adjusting cosmological and physical parameters one can construct a universe without weak interactions that still has nucleosynthesis, matter domination, structure formation, stars, and supernovas (Harnik, Kribs & Perez 2006).

Second, the “fine tuning” may be less exact than it appears. (Livio et al. 1989), for example, note that a changed Hoyle resonance does not necessarily change carbon production enormously and that the scale of what counts as fine tuning in this case is subjective. Many parameters can change by an order of magnitude: is this large? In many cases the limits mentioned above are firm if only one parameter is varied, but far more yielding if two or more parameters change together (Stenger 2011)[21].

There has been some effort to measure the sensitivity of physics to changes in fundamental constants. For example, (Ellis et al 1986; Barbieri & Giudice 1988) measure it by “unnaturalness” $\Delta \equiv \left|\frac{\partial \ln c}{\partial \ln a}\right|$ where c is some model parameter affected by parameters of a more fundamental theory. If Δ has a large value the world will be strongly affected by small changes in a, hinting that fine tuning of some kind is needed. Another approach is to view something as “unnatural if it has parameters very close to special values that are not explained by the symmetries of the theory.” (Hall & Nomura 2008). This in particular relevant if continuous changes of the parameters cause drastic differences in behaviour, such as the existence of complex structures, even if the underlying

[19] They are however not *unknown* among other chemicals. For example, silicon, gallium, germanium, antimony, bismuth, plutonium, and acetic acid expand when they freeze.

[20] An interesting fictional but physics-based example is the carefully worked out five-dimensional world in Greg Egan’s novel *Diaspora* (1997). Several papers on the dimensionality of space-time have noted the instability of electron orbits in higher dimensions, but Egan shows that the solution of the Schrödinger equation can be interpreted as a different kind of atom (one that has electrons residing within the nucleus, with orbitals “poking out”) that may have potential for chemistry. That this possibility had been overlooked in the literature suggests that we may indeed be parochial in declaring different kinds of physics unable to sustain life or other complex phenomena.

[21] However, see (Barnes 2012) for a detailed critique of Stenger.

theory does not do anything drastic near these boundaries: finding the parameters close to such a boundary should give rise to suspicion.

Third, if the universe is large – either because the inflation theory is correct and distant regions have other settings of many physical constants, or because there is a multiverse with separate universes with different physical laws – then there is no reason to be surprised about these fine tunings (Tegmark 1998; Hall & Nomura 2008). Observers are only found where the laws of physics are benign.

Whether fine-tuning can be used as evidence for large or ensemble universes is controversial. String theory and eternal inflation naturally produces large universes (Schellekens 2013), so some thinkers regard them as lending credence to anthropic principle (Susskind 2007). Others argue that anthropic principles cannot produce any falsifiable predictions, although ensemble universe theories can (Smolin 2007).

Boltzman brains

In the *truly* long run, if the universe persists, then even exceedingly unlikely processes will occur. This is also true if the universe is spatially infinite. Boltzmann originally proposed (Boltzmann 1895) that the entire universe could be a low-probability low-entropy fluctuation in a high-entropy world.

Smaller fluctuations are exponentially more likely. That means that occasionally vacuum fluctuations will not just randomly produce particle-antiparticle pairs, but atoms, molecules, and entire organisms that briefly persist. This includes randomly generated observers that can have brains in arbitrary states – the “Boltzmann brains”.

Boltzmann brains are a problem since they are anomalous observers. Their memories are arbitrary and have no causal connection to the outside universe. But if the universe persists long enough or is infinite, nearly all observers will be anomalous rather than normal. Hence, we should assign a high probability to being Boltzmann brains ourselves. (Barrow & Tipler 1986; Albrecht & Sorbo 2004; Bousso & Freivogel 2007)

This seems to lead to radical sceptical problems. Boltzmann brains do not have accurate information about the outside world, and believing one to be a Boltzmann brain does seem to be self-defeating (since their understanding of physics is likely to be totally wrong). They also tend to have incoherent experiences of the world, so while selection effects might give you coherent past memories if your experiences right now do not turn incoherent you have a reason to doubt being an anomalous observer (Davenport & Olum 2010). In fact, given that anomalous observers can experience *anything*, the hypothesis that you are one is not supported very strongly by your evidence (Huemer 2015).

People have tried to show that the number of normal observers outweigh the anomalous observers.

One way out of the problem is to argue that the universe will be short-lived and is not of infinite spatial extent. Page estimated that if the universe persists less than 10^{51} years the paradox does not occur (Page 2005), but with cosmic acceleration this might have to happen within 20 billion years (Page 2008).

Others have argued that the far-future states are not amenable to produce Boltzmann brains, e.g. (Gott 2008; Boddy, Carroll & Pollack 2015). The problem may be that they can occur both because of quantum effects and thermodynamic miracles in spatially big universes, making them hard to rule out (Davenport & Olum 2010).

Much of the debate hinges on the measure problem: how do we use probability across infinite spacetimes or across multiverse ensembles? Are Boltzmann brains an argument against certain theories or measures? (Linde & Noorbala 2010; de Simone et al. 2010)

This is a case where a "big" universe causes more problems than it solves: in the fine tuning case observer selection effects makes fine tuning trivial if the total universe is large and varied enough, but in the Boltzmann brain case the larger the universe is, the more peculiar it is that we are not Boltzmann brains.

Conclusion

Observer selection effects are common, especially when considering the big questions facing humanity. I think there are two main lessons to draw from this domain:

The first is that observation effects can produce nontrivial biases in data that we may have to control for, or produce real biases in the observed world. The more strong correlations are with our existence as observers, the more bizarre the biases.

The second is that anthropic reasoning is often a decent prior when there is no data or merely the single data point of our existence. However, when more data arrives it tends to overwhelm the theory quickly[22]. This is largely a good thing: anthropic reasoning is notoriously controversial, confusing and easy make mistakes in. But sometimes we do not have any data and then it is the best we can do.

[22] But not always: see the example in (Verendel & Häggström 2015).

References

***=Outstanding

**=Recommended

*=Relevant

Albrecht, A., & Sorbo, L. (2004). Can the universe afford inflation?. *Physical Review D*, 70(6), 063528.

Armstrong, S., & Sandberg, A. (2013). Eternity in six hours: Intergalactic spreading of intelligent life and sharpening the Fermi paradox. *Acta Astronautica*, 89, 1-13.

Barbieri, R., & Giudice, G. F. (1988). Upper bounds on supersymmetric particle masses. *Nuclear Physics B*, 306(1), 63-76.

* Barnes, L. A. (2012). The fine-tuning of the universe for intelligent life. *Publications of the Astronomical Society of Australia*, 29(4), 529-564.

Barrow, J. D. (1983). Anthropic definitions. *Quarterly Journal of the Royal Astronomical Society*, 24, 146.

*** Barrow, J. D. & Tipler, F. J. 1986. *The Anthropic Cosmological Principle*. Oxford: Clarendon Press.

Bell, E. A., Boehnke, P., Harrison, T. M., & Mao, W. L. (2015). Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proceedings of the National Academy of Sciences*, 201517557.

* Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review*, 386-398.

Boddy, K. K., Carroll, S. M., & Pollack, J. (2015). Why Boltzmann Brains Don't Fluctuate Into Existence From the De Sitter Vacuum. arXiv preprint arXiv:1505.02780.

Boltzmann, L. (1895). On certain questions of the theory of gases. *Nature*, 51, 413-415.

* Bostrom, N. (1998). The Doomsday Argument: a Literature Review. Online at http://anthropic-principle.com/preprints/lit/index.html

Bostrom, N. (2001). Cars in the next lane really do go faster. *Plus Magazine*.

Bostrom, N. (2001b). The Doomsday Argument Adam & Eve, UN++, and Quantum Joe. Synthese, 127(3), 359-387

** Bostrom, N. (2003). Are we living in a computer simulation?. *The Philosophical Quarterly*, *53*(211), 243-255.

* Bostrom, N. (2003b). Astronomical waste: The opportunity cost of delayed technological development. *Utilitas*, 15(03), 308-314.

Bostrom, N., & Ćirković, M. M. (2003c). The doomsday argument and the self-indication assumption: reply to Olum. *The Philosophical Quarterly*, 53(210), 83-91.

Bostrom, N. (2008). Where are they? Why I hope the search for extraterrestrial life finds nothing. *Technology Review*, 72-78.

*** Bostrom, N. (2010). *Anthropic bias: Observation selection effects in science and philosophy*. Routledge.

* Brin, G. D. (1983). The great silence-The controversy concerning extraterrestrial intelligent life. *Quarterly Journal of the Royal Astronomical Society*, 24, 283-309.

Bousso, R., & Freivogel, B. (2007). A paradox in the global description of the multiverse. *Journal of High Energy Physics*, 2007(06), 018.

Bowring, S. A., Erwin, D. H., & Isozaki, Y. (1999). The tempo of mass extinction and recovery: The end-Permian example. *Proceedings of the National Academy of Sciences*, 96(16), 8827-8828.

Carter, B. (1974). Large number coincidences and the anthropic principle in cosmology. In *Confrontation of cosmological theories with observational data* (pp. 291-298). Springer Netherlands.

** Ćirković, M. M., Sandberg, A., & Bostrom, N. (2010). Anthropic shadow: Observation selection effects and human extinction risks. *Risk analysis*, *30*(10), 1495-1506.

Dar, A., De Rújula, A., & Heinz, U. (1999). Will relativistic heavy-ion colliders destroy our planet?. *Physics Letters B*, 470(1), 142-148.

Davenport, M., & Olum, K. D. (2010). Are there Boltzmann brains in the vacuum?. arXiv preprint arXiv:1008.0808.

Davies, P. C. W. (1972). Time variation of the coupling constants. *Journal of Physics A: General Physics*, 5(8), 1296.

Davies, P. C. (1982). *The accidental universe*. Cambridge University Press.

Dicke, R. H. (1957). Gravitation without a principle of equivalence. *Reviews of Modern Physics*, 29(3), 363.

Dicke, R. H. (1961). Dirac's Cosmology and Mach's Principle. *Nature*, 192, 440-441.

Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis*, 143-147.

Ellis, J., Enqvist, K., Nanopoulos, D. V., & Zwirner, F. (1986). Observables in low-energy superstring models. Modern Physics Letters A, 1(01), 57-69.

Ellis, J., Giudice, G., Mangano, M., Tkachev, I., & Wiedemann, U. (2008). Review of the safety of LHC collisions. *Journal of Physics G: Nuclear and Particle Physics*, 35(11), 115004.

Eom, Y. H., & Jo, H. H. (2014). Generalized friendship paradox in complex networks: The case of scientific collaborattion. *Scientific reports*, *4*, 4603 (2014).

Epelbaum, E., Krebs, H., Lähde, T. A., Lee, D., & Meißner, U. G. (2013). Viability of carbon-based life as a function of the light quark mass. *Physical review letters*, 110(11), 112502.

Feld, S. L. (1991). Why your friends have more friends than you do. *American Journal of Sociology*, 1464-1477.

Giddings, S. B., & Mangano, M. L. (2008). Astrophysical implications of hypothetical stable TeV-scale black holes. *Physical Review D*, 78(3), 035009.

** Gott, J. R. (1993). Implications of the Copernican principle for our future prospects. *Nature*, 363(6427), 315-319.

Gott III, J. R. (2008). Boltzmann Brains--I'd Rather See Than Be One. arXiv preprint arXiv:0802.0233.

* Grace, C. (2010). *Anthropic Reasoning in the Great Filter*. Bachelor thesis, Australian National University.

Grund, T.U. (2014).“Why Your Friends Are More Important and Special Than You Think.” *Sociological Science* 1: 128–140.

Hall, L. J., & Nomura, Y. (2008). Evidence for the Multiverse in the Standard Model and Beyond. Physical Review D, 78(3), 035001.

** Hanson, R. (1998). The great filter-are we almost past it?. Preprint available at http://hanson.gmu.edu/greatfilter.html

Harnik, R., Kribs, G. D., & Perez, G. (2006). A universe without weak interactions. Physical Review D, 74(3), 035006.

Hellman, M. E. (2008). Risk analysis of nuclear deterrence. *The Bent of Tau Beta Pi*, Spring, (14-22).

Hodas, N. O., Kooti, F., & Lerman, K. (2013). Friendship paradox redux: Your friends are more interesting than you. *arXiv preprint arXiv:1304.3480*.

Höhle, M., & Held, L. (2006). *Bayesian Estimation of the Size of a Population* (No. 499). Discussion paper//Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München.

Huemer, M. (2015) Serious theories and skeptical theories: Why you are probably not a brain in a vat. *Philosophical Studies*, 1-22.

Jaffe, R. L., Jenkins, A., & Kimchi, I. (2009). Quark masses: An environmental impact statement. *Physical Review D*, 79(6), 065014.

Kragh, H. (2010). An anthropic myth: Fred Hoyle’s carbon-12 resonance level. *Archive for history of exact sciences*, 64(6), 721-751.

Leslie, J. (1990). Is the end of the world nigh?. *The Philosophical Quarterly*, 65-72.

* Leslie, J. (1998). *The End of the World: the science and ethics of human extinction*. Psychology Press.

Linde, A., & Noorbala, M. (2010). Measure problem for eternal and non-eternal inflation. *Journal of Cosmology and Astroparticle Physics*, 2010(09), 008.

Lineweaver, C. H., & Egan, C. A. (2012). The initial low gravitational entropy of the universe as the origin of design in nature. In *Origin (s) of Design in Nature* (pp. 3-16). Springer Netherlands.

Livio, M., Hollowell, D., Weiss, A., & Truran, J. W. (1989). The anthropic significance of the existence of an excited state of 12C. *Nature* 281-284.

Lynden-Bell, R. M., Morris, S. C., Barrow, J. D., Finney, J. L., & Harper, C. (Eds.). (2010). *Water and life: The unique properties of H2O* (Vol. 20). CRC Press.

Mangel, M., & Samaniego, F. J. (1984). Abraham Wald's work on aircraft survivability. *Journal of the American Statistical Association*, *79*(386), 259-267.

Moravec, H. (1988). *Mind children: The future of robot and human intelligence*. Harvard University Press.

Newman, W. I., & Sagan, C. (1981). Galactic civilizations: Population dynamics and interstellar diffusion. *Icarus*, 46(3), 293-327.

Olum, K. D. (2002). The doomsday argument and the number of possible observers. *The Philosophical* Quarterly, 52(207), 164-184.

Ord, T., Hillerbrand, R., & Sandberg, A. (2010). Probing the improbable: methodological challenges for risks with low probabilities and high stakes. *Journal of Risk Research*, 13(2), 191-205.

Ord, T. (2014) Using a fund manager? You'd get the same results at a casino. *The Conversation* August 7. 2014

Page, D. N. (2005). The Lifetime of the Universe. *J. Korean Phys. Soc*., 49(hep-th/0510003), 711-714.

Page, D. N. (2008). Is our universe likely to decay within 20 billion years?. *Physical Review D*, 78(6), 063535.

Rees, M. J. (2000). Just six numbers: The deep forces that shape the universe. Basic Books.

* Robock, A., Oman, L., & Stenchikov, G. L. (2007). Nuclear winter revisited with a modern climate model and current nuclear arsenals: Still catastrophic consequences. *Journal of Geophysical Research: Atmospheres* (1984–2012), 112(D13).

Ruggles, R., & Brodie, H. (1947). An empirical approach to economic intelligence in World War II. *Journal of the American Statistical Association,42*(237), 72-91.

Sagan, S. D. (1995). *Limits of safety: Organizations, accidents, and nuclear weapons* (No. CONF-950264--). American Association for the Advancement of Science, Washington, DC (United States).

Sahney, S., & Benton, M. J. (2008). Recovery from the most profound mass extinction of all time. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1636), 759-765.

Schellekens, A. N. (2013). Life at the interface of particle physics and string theory. *Reviews of Modern Physics*, 85(4), 1491.

* Schlosser, E. (2013). *Command and control: Nuclear weapons, the Damascus accident, and the illusion of safety*. Penguin.

De Simone, A., Guth, A. H., Linde, A., Noorbala, M., Salem, M. P., & Vilenkin, A. (2010). Boltzmann brains and the scale-factor cutoff measure of the multiverse. *Physical Review D*, 82(6), 063520.

Smolin, L. (2007). Scientific alternatives to the anthropic principle. In *Universe or multiverse*, ed. Carr, B. Cambridge University Press. pp. 323-366.

Steinhart, E. (2010). Theological implications of the simulation argument. *Ars Disputandi*, 10(1), 23-37

Stenger, V. (2007). The Anthropic Principle. In *The Encyclopedia of Nonbelief*. Tom Flynn, ed., Prometheus Books 2007, pp. 65-70.

Stenger, V. J. (2011). *The fallacy of fine-tuning: Why the universe is not designed for us*. Prometheus Books.

Susskind, L. (2007). The anthropic landscape of string theory. In *Universe or multiverse*, ed. Carr, B. Cambridge University Press. pp. 247-266.

Tegmark, M. (1997). On the dimensionality of spacetime. *Classical and Quantum Gravity*, 14(4), L69.

** Tegmark, M. (1998). Is "the theory of everything" merely the ultimate ensemble theory?. *Annals of Physics*, 270(1), 1-51.

Tegmark, M., & Rees, M. J. (1998). Why is the cosmic microwave background fluctuation level 10−5?. The Astrophysical Journal, 499(2), 526.

** Tegmark, M., & Bostrom, N. (2005). Astrophysics: Is a doomsday catastrophe likely?. *Nature*, 438(7069), 754-754.

Tegmark, M., Aguirre, A., Rees, M. J., & Wilczek, F. (2006). Dimensionless constants, cosmology, and other dark matters. *Physical Review D*, 73(2), 023505.

Tipler, F. J. (1988, January). The anthropic principle: a primer for philosophers. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (pp. 27-48). Philosophy of Science Association.

Verendel, V., & Häggström, O. (2015). Fermi's paradox, extraterrestrial life and the future of humanity: a Bayesian analysis. arXiv preprint arXiv:1510.08684.

** Wainer, H., Palmer, S., & Bradlow, E. T. (1998). A selection of selection anomalies. *Chance*, 11(2), 3-7.

Wald, A. (1980). *A Reprint of A Method of Estimating Plane Vulnerability Based on Damage of Survivors* (No. CRC-432). CENTER FOR NAVAL ANALYSES ALEXANDRIA VA OPERATIONS EVALUATION GROUP.

Wallace, A. R. (1903). *Man's place in the universe: a study of the results of scientific research in relation to the unity or plurality of worlds*, New York, McClure, Phillips & co.

Waltham, D. (2011). Testing anthropic selection: A climate change example. *Astrobiology*, 11(2), 105-114.

Ward, P. & Brownlee, D. (2000). *Rare Earth: Why Complex Life is Uncommon in the Universe*, Copernicus Books

Webb, S. (2002). *If the universe is teeming with aliens... where is everybody?: fifty solutions to the Fermi paradox and the problem of extraterrestrial life* (Vol. 101). Springer Science & Business Media.

Weinberg, S. (1987). Anthropic bound on the cosmological constant. *Physical Review Letters*, 59(22), 2607.