Counterexamples to the Anthropic Principle

Shamik Dasgupta
Draft of August 2015

The Anthropic Principle states that our existence is a selection effect. Suppose some process can produce a number of outcomes $O_1 \ldots O_n$, but only O_i is consistent with our existence. And suppose the objective probability that the process would produce O_i is greater if a certain hypothesis H were true, than were it false:

$$P(O_i \mid H) > P(O_i \mid -H)$$

Does observing O_i confirm H? One might think so, on the grounds that observations confirm those hypotheses on which they are more probable. But in this case, our being in a position to make any observations at all guarantees that we will observe O_i. So, according to the Anthropic Principle, the fact that we observe O_i is unsurprising, and should teach us nothing. As Carter (1974) put it, "what we can expect to observe must be restricted by the conditions necessary for our presence as observers". Formally, the Anthropic Principle states that we must conditionalize our observations on our existence. In this case, our existence implies O_i, so

$$P(O_i \mid \text{I exist}) = 1$$

and hence

$$P(O_i \mid H \ \& \ \text{I exist}) = 1 = P(O_i \mid -H \ \& \ \text{I exist}).$$

Conditional on the fact that I exist, O_i is not more likely given H than given -H; hence O_i does not confirm H after all.[1]

The Anthropic Principle has recently become a trusted part of the atheist's tool-kit, and is used frequently to disarm the fine-tuning argument. A number of physical constants, we are told, must be set within a narrow range for life to be possible. According to the fine-tuning argument, it is vastly more probable that those constants were set like that if they were designed by an agent intent on creating life (God), than if they were set by chance. Let "fine-tuned" be the observation that the constants are finely-tuned; let "design" be the hypothesis that they were set by design; and let "chance" be the hypothesis that they were set by chance. Then the fine-tuning argument states that

$$P(\text{fine-tuned} \mid \text{design}) > P(\text{fine-tuned} \mid \text{chance})$$

[1] This sometimes called the *weak* Anthropic Principle. Unfortunately, this title has been used to describe a bewildering array of different ideas; my focus here is just on the idea described in the text.

and so our observing that the constants are fine-tuned is evidence for design over chance. But according to the Anthropic Principle, we must conditionalize on our existence. Since our existence implies that the constants must be fine-tuned, we get

P(fine-tuned | design & I exist) = 1 = P(fine-tuned | chance & I exist)

and so observing that the constants are fine-tuned lends no weight to design after all.[2]

But the Anthropic Principle is incorrect. Leslie (1989) tried to demonstrate this with his firing-squad example. Suppose you are sentenced to death and face 100 trained marks-persons. You hear the shots fired, but instead of blacking out, you open your eyes to see 100 bullet holes in the wall behind you. Intuitively, this happy outcome is evidence that there was a conspiracy to save you. After all, your survival is much more likely if there was a conspiracy than if they had all missed by chance. But according to the Anthropic Principle, your survival teaches you nothing about anything, for the fact that you are in a position to make any observations at all guarantees that you survived.

I am convinced by this example, but fans of the Anthropic Principle are not. They are happy to bite the bullet, treating it as a recherché example of little importance. Why? Perhaps because in the firing-squad case, the evidence is *one's continued survival*, and real scientific evidence is rarely (if ever) like this. Insofar as our aim is to model good scientific reasoning, perhaps it is not a serious problem that the Anthropic Principle gives strange results in this fictional case. Other refutations of the Anthropic Principle are similar: they describe far-fetched cases in which, intuitively, *one's waking up*, or *one's being alive*, is evidence for some hypothesis.[3] We can expect fans of the Anthropic Principle to be unmoved by these cases too, and for the same reason.

A better refutation would show that the Anthropic Principle is contrary to good scientific reasoning. We have seen that, for any putative piece of evidence E that satisfies the equation

(*) P(E | I exist) = 1,

the Anthropic Principle says that E is no evidence for anything. I will argue that all manner of established scientific evidence satisfy (*). So the problem with Anthropic Principle is not (just) that it goes wrong in the firing-squad example, but that it is *inconsistent with good scientific practice*.

---

[2] For examples of this use of the Anthropic Principle, see Penrose (1989), Sober (2004), and Dawkins (2006).

[3] See, for example, Zuboff's (1990) example of waking up in the hotel room, or Kotzen's (2014) example of figuring out whether your genetic origins were safe.

Let me start with some warm-up examples that illustrate the central idea, before developing more complex examples that are in some respects better. If my father had never been born, I would not exist. My mother may have had children with someone else, but she would not have had *me*. For *me* to exist, it was a necessary condition that my biological parents existed to conceive me. The same goes for them: a necessary condition for their existence was that their parents existed. And so on down the line. Thus, for any ancestor of mine *x*, my existence implies *x*'s existence, and so

P(*x* exists | I exist) = 1

So, according to Anthropic Principle, learning that *x* exists can teach me nothing about anything. This is wrong. A hypothesis about medieval politics may hang on the existence of a hypothetical monarch, so that discovering that the monarch really existed confirms the hypothesis. But not, says the Anthropic Principle, if one happens to be descended from the monarch. My maternal grandfather traced his line back to the Plantagenet kings. According to the Anthropic Principle, this alone is enough to rule out my career as a medieval historian.

A more realisitc example comes from paleoanthropology. I am a Homo *sapiens*. This is not just an accident; it is not as if I could have been a sea cucumber instead. I am necessarily a Homo *sapiens*, and so a necessary condition on my existence is that Homo *sapiens*, the species, exists. Now, Homo *sapiens* evolved from Homo *habilis*. If Homo *habilis* had never evolved, nor would Homo *sapiens*. Other similar species qualitatively like ours may have evolved, but it would not have been *our very species*, for species, it seems, are like individual people insofar as their existence depends on that of their predecessors. Thus, a necessary condition for the existence of Homo *sapiens* is the existence of Homo *habilis*. Likewise, a necessary condition on the existence of the latter is the existence of *its* predecessors, such as Australopithecus *afarensis*. Thus,

P(A. *afarensis* existed | I exist) = 1.

According to the Anthropic Principle, then, the discovery of A. *afarensis* can teach us nothing about anything. That would be news to paleoanthropologists, for whom the discovery was an event of great significance. Until then, it was unknown whether bipedalism evolved before or after the increase in brain size. But Lucy, the first A. *afarensis* discovered, was a small-brained biped, confirming the hypothesis that bipedalism came first.

Once you see the pattern, the examples quickly multiply. If this planet didn't exist, I would not exist: a life-form would not be *me* if it did not bud from this very planet. Sure, a qualitatively similar life-form could have sprouted elsewhere in the universe instead, but it would not be *me*, it would just look like me. So my existence implies the existence of this planet.[4] Similarly, the existence of this planet implies the existence of the Milky Way. A qualitatively similar planet that

[4] The point here is not that *life* would not exist without this very planet. That is surely false. The point is that *I* would not exist had a life-form not emerged from this very planet.

formed from distinct matter in some far flung part of the universe would not be *this very planet*; it would just look very much like this planet. So, my existence implies the existence of the Milky Way. According to the Anthropic Principle, then, discovering that the Milky Way exists can teach us nothing about anything.[5]

These warm-up examples illustrate the main idea: facts about the past can be evidence, even if they are necessary conditions for our existence. One might object that my existence does not *logically* imply the existence of my parents, or of our planet: it is logically possible for me to be the only thing that exists. This is true, but so what? Our existence does not *logically* imply that the constants are fine-tuned either, yet when one uses the Anthropic Principle to refute the fine-tuning argument, one insists that our existence implies that the constants are fine-tuned. Clearly, the notion of implication used by the Anthropic Principle is not logical implication, but what is sometimes called "metaphysical" implication. I am just following suit.[6]

The examples assume "origin essentialism", the idea that ordinary objects could not have had radically different origins. This yields the thought that for a life-form to be *me*, it must come from certain parents; that for a rock to be *this very planet*, it must originate from some particular quantity of matter. I will not defend origin essentialism here, in part because it is very popular these days, but also because it would be interesting enough to show that one can use the Anthropic Principle only if one rejects origin essentialism. Not for the first time, a principle wielded in the name of science would be found to rest on weighty metaphysical presuppositions.

A more pressing objection is that, when we confirm some hypothesis about medieval politics, our evidence is never the proposition *that a certain person existed*, but the stronger proposition that he existed *and led such-and-such a life*—for example, that he won a battle. Though I could not have existed without my ancestor, I could still have existed if my ancestor lost instead. So my existence does not imply the evidence, and so the evidence does not satisfy equation (*) after all. Similarly, one might object that the archeological evidence was not the weak proposition that A. *afarensis* existed, but the stronger proposition that it existed *and was a small brained bipedal*. So, one might object that my existence it does not imply this evidence; so, again, the evidence does not satisfy equation (*).

More generally, the objection is that real scientific evidence is rarely (if ever) that something exists, but that it exists *and has a certain property*. And while my existence might imply that those things exist, it does not imply that they have those properties.

---

[5] When defending the Anthropic Principle against the firing-squad example, Sober (2004) pointed out that the principle allows that my survival would be evidence *to a third party* that there was a conspiracy, since their existence does not imply that I survived. Clearly, this kind of response is ineffective against these last two examples: there are no third parties to appeal to.

[6] Relatedly, one might complain that we should use a probability function that does not always assign $P(X \mid Y) = 1$, when Y implies X. Perhaps so. But when one uses the Anthropic Principle to dismantle the fine-turning argument, one uses a probability function with this property. Again, I am just following suit.

I suspect that (at least one of) the examples above can be defended.[7] But no matter, for we can develop more complex examples that side-step the objection. In these examples, my existence implies the existence of other things *and their having certain properties*, and their having those properties is undeniable scientific evidence.

To show this, start again with my ancestors. We know that my existence implies their existence; the objection was that it does not imply much about what their life was like. But my existence does imply *something* about their life. For me to exist, it is not just that my parents had to exist; they also had to conceive within some window of time. If they had waited a month—a few days even—the specific sperm and egg that I was in fact conceived from could not have fused, and the child they would have conceived would not have been *me*. Perhaps they did not have to conceive at the *very* moment they did—perhaps it is possible, though extraordinarily unlikely, for them to conceive me if they had waited an hour, or perhaps even a day. Still, there is some interval of time within which they *had* to conceive for me to exist. Let *t* be that interval of time, and let E be the proposition that they conceived within *t*. My existence implies E; hence E satisfies equation (*); hence, according to the Anthropic Principle, E can teach *me* nothing about anything.

And that is clearly wrong. Suppose that my father is currently suspected of having being a spy back in the 1970s. And suppose that the case against him consists of activities in Russia that occurred throughout the interval *t*. He can therefore establish his innocence by showing that at *some* point in *t* he was not spying. With luck, a video recording of my parents, in the act of conception, is found (apologies for the graphic overtones, but believe me this is more awkward for me than for you). It has no specific time-stamp, but is known to have been recorded at some time during *t*. Thus, my father can establish E to the jury, and on that basis he is exonerated. But according to the AP, I should still doubt my father's innocence. I know that the evidence demonstrates his innocence to everyone else. But according to the Anthropic Principle, it demonstrates nothing *to me*. This is clearly absurd.

Similar, more scientific examples, are forthcoming. We saw that my existence implies the existence of planet Earth; the objection was that it does not imply anything about what this planet is like. But that is false. Our planet was formed roughly 4.5 billion years ago. For it to exist, it had to have been formed during a certain window of time. If a qualitatively similar planet was formed billions of years later instead, in a very different stage of the Milky Way's life, it would not be *this planet*. Thus, my existence doesn't just imply that our planet exists; it also implies that it was formed roughly 4.5 billion years ago. Hence, according to the Anthropic Principle, discovering that our planet was formed roughly 4.5 billion years ago can teach us nothing about

---

[7] In the case of A. *afarensis*, let us grant that the evidence was that A. *afarensis* existed and was a small brained biped. The trouble with the objection is that it is not clear that A. *afarensis* could have been anything other than a small brained, bipedal species. That is, one might think that being a small brained biped is essential to A. *afarensis*, so that if my existence implies the weaker proposition that A. *afarensis* exists, then it also implies the stronger proposition that A *afarensis* existed and was a small brained biped.

anything. Think what you like about firing squads, this is obviously an intolerable result. The fact that our planet is roughly that old is important scientific evidence in favor of the theory of evolution by natural selection. Thus, the Anthropic Principle is contrary to good scientific method.

This last example reveals a deep irony in recent atheism. I am an atheist, so I would like to show that the fine-tuning argument is unsound. And I would like to do so on scientific grounds, by showing that there are scientific explanations of our existence. My guess is that Penrose (1989), Sober (2004), and Dawkins (2006) feel the same way. But if one appeals to the Anthropic Principle, as they do, one appeals to something that is contrary to the very scientific standards used to affirm the theory of evolution by natural selection—itself a central piece of the scientific explanation of our existence. Our defense of atheism must come from elsewhere.

Bibliography

Carter, Brandon (1974). "Large Number Coincidences and the Anthropic Principle in Cosmology". In *Confrontation of Cosmological Theories with Observational Data*, edited by M. S. Longair, pp. 291-8. Dordrecht: Reidel.

Dawkins, Richard (2006). *The God Delusion*. Bantam Press.

Kotzen, Matthew (2012). "Selection Biases in Likelihood Arguments". *The British Journal for the Philosophy of Science* 63: 825-839.

Leslie, John (1989). *Universes*. London: Routledge.

Penrose, Roger (1989). *The Emperor's New Mind*. Oxford: Oxford University Press.

Sober, Elliot (2004). "The Design Argument". In *The Blackwell Guide to the Philosophy of Religion*, edited by William E. Mann, pp. 117-147.

Zuboff, Arnold (1990). "One Self: The Logic of Experience" *Inquiry* 33: 39-68.