

Eris in Retrograde

 hivewired.wordpress.com/2019/01/14/eris-in-retrograde/

Saede Riordan

January 14, 2019

Epistemic Status: Firmly supported theories and personal observations

During the summer of 2017, I started work on the Origin sequence, a project I hoped would move the rationality community in a positive direction, encourage more people to participate, and create something memetically contagious that would spread throughout the population. In the ideal case, this would snowball into a spiral of enlightenment that totally remakes society. It was...kind of ridiculous, “this project will save the world!” yeah okay past me, calm down a bit.

I was definitely biting off more than I could chew. Despite mostly still agreeing with the goals of the project, there’s little I would endorse in my past approach. I didn’t really understand what I was getting into when I started, how massive and complicated a project it would become, or how deeply embedded the associated ethical interpretations were in the psyche. Fatally I failed to appreciate how difficult it would be to steer and modify an accelerating and expanding system.

So now we’re a year Post-Origin, and the beginning of the new year seems like a good time to write a postmortem and discuss the future. We’ll go over what the issues were, what went wrong, and finally a discussion of why these things are difficult.

Foundational Quicksand

Origin was probably doomed from the start. It was very grand and very optimistic, and basically completely failed to take into account the existence of things like ‘reality’ or ‘other people’ during the initial planning. The long-term goal of Origin was to remake society in the image of rationality, to modify the rationality memeplex to be more virulent and compelling so that it spread throughout the population and raised the sanity waterline. There was a baked in assumption here that becoming more rational always makes you a better person. I figured that the knock-on effects of this would be a wonderful and better world with people who are smarter, kinder, and more reasonable, with no downsides. That goal is on the level of “construct a nuclear weapon in your garage” when you really start to break down the nuts and bolts. If I didn’t know better I might say we were playing with fire. But really? We were at the level of playing with flint and wet sticks. We would have been lucky to generate sparks, much less flame, much less actually burn anyone, intentionally or otherwise.

Our goals were naive, literally based on an April Fools post by Eliezer, and I no longer see them as corresponding to reality in a way that makes them valuable to pursue. Moreover, there’s an argument it was actually dangerous. If you taught rationality techniques to Donald Trump (and he managed to actually internalize them), would that be good for the world or bad? Depending on how good at instilling morality you think the rationality memeplex is, it might be good, but it seems to me that it’s much more

likely it would just make him better at being bad.

What eventually did in the project was tripping over its own premature implementation. We didn't have a coherent idea of what we were trying to do, but we rushed into doing it anyway. When I presented ideas to spitball off they were uncritically accepted and rushed into practice. I wanted to create a memplex, but many of the other people involved in the project acted as if the memplex I sought to create already existed. Eventually, the rushed implementation ran into the lack of foundation, got lost in bikeshedding, fell over, and died.

I care about this because I care about the world and I don't want to become captured by a goal that won't contribute to saving the world. There's too much at stake to waste effort and motion on something that won't work. Better to cast off the sunk cost and begin anew.

Time for some Thrilling Memetics

The failures and impossibilities of Origin's goals can in some sense be extracted from a description of them alone. I was pretty naive in my understanding of memetic forces at the time, to think I could simply find the right aesthetics and then people would eat up the rationality memes because they seemed cool.

The basic factor that I was almost willfully ignoring is that the part of a memplex that makes it spread tends to be the particular meme saying "spread these associated memes in the memplex, don't let these competing memes in" or in other words, "shun the heretics and nonbelievers, you must believe in god this much to enter the kingdom of heaven."

There's not some hidden memetic master technique that makes people fall for religion, the trick is "spread this meme or else" and there's not really much more to it than that. In a sense, I was both over and under-thinking the problem. I sincerely thought I could make something memetically virulent without the "spread this meme or else" subcomponent, missing that "spread this meme or else" is one of the things creating that virulence in the first place.

What else is? A sense that a particular view is not only correct but moral, that it's the righteous choice to make to be a follower of God. That everyone making a different choice is not only wrong but immoral and probably evil.

The principal memetic subcomponents to making a supermeme as far as I can tell at this point, one that will spread with the virulence of the Spanish flu through an anti-vaxxer convention, are the following sub-memes:

- Spread this Memplex.
- If you don't spread this memplex, then Bad Thing will happen to you.
- This memplex is correct, opposing memplexes are incorrect, doubting this is bad and makes you an idiot.

- This meme is the most moral and will lead to the best outcomes if you use it.
- This meme is powerful and will give you an edge by conferring an advantage to you over your competitors.

Now, none of these statements need to be true, they just need to be present. The fact that their presence is more important than whether they are true statements about the world is kind of a problem. I was unsure if I wanted to even post the specific sub-memes for this out of concern that someone with fewer scruples might take those things and make a weaponized memeplex. But even with all those parts, most cults still fail. This isn't an easy thing to get right *even if* you're willing to disregard all scruples and be completely ruthless and manipulative. And if you're handicapping yourself by not being completely evil? In most cases you can forget about it.

We're not in the old world where starting a religion is somewhat viable, the internet has pushed us into realms unknown. The Tower is a harsh environment for any memeplex.

“Happiness and meaning—sometimes they overlap, sometimes you must choose. I don’t have the answer, there is no answer, all I can do is warn you about the trap by which you obtain neither. Even if you’re sign-me-up-for-the-Orgasmatron all in with Team Experiencing Self, the id is too myopic to be any good at long-term hedonism. The unchecked id would have left us cavemen, samizdat & chill wouldn’t even be on the table. Conversely, even if you’re polishing trophies for Team Remembering Self, the default superego is an incoherent mess, infected with millions of selfish, **MALIGNANTLY USELESS** memes that have no interest in your happiness, care not for the coherence of your autobiography, and will drive you to madness rather than let you winnow them away.

The key word is default. We all have some degree of protection, either through physical isolation or memetic immunity, “Mom says not to trust strangers who say they have candy.” But most of us fall short of contact precautions. And in that case, we are ruled by probability—by Moloch, by Nyx, by Nature, the only force that God fears. Why else would He confuse mankind’s language? Why would He demand obedience to 613 commandments? Circumcision? What was Judaism, with rabbinical prohibition against interfaith marriage or proselytization, except God’s attempt to create a religion that would not spread? It failed, as it always does. Autotune and Manifest Destiny. The house always wins at the second law of thermodynamics.

With free flow of information, how can any belief system hold? All belief systems rest on axioms, if you grant equal footing to a contradictory axiom, the belief system collapses. I suppose I’m that guy claiming that atheists invent a God—not an interventionist God, nor a fuzzy deism, but a set of unprovable principles that determine right and wrong and to which one must atone. Don’t give me that humanism bullshit. When someone slaps your hypothetical girlfriend’s ass in the proverbial club, what does humanism say you should do? At least toxic masculinity has an answer. Humanism is a motte and bailey, a set of milquetoast ideals which provide no guidance in day to day life and so leave you passive (“Hey, man—first principles!”) or, more likely, vulnerable to whatever crypto-ideology is most virulent. If you do not have a code of conduct, one will be provided for you.

With free flow of information—a suppressed memetic immune system, a hypothetical Tower of Babel—it is statistically inevitable that every meme will attain its most infectious form. There are countless ways to make an idea more or less palatable, but the first step is always the same, a single amino acid substitution, a lingering desire affixed to every thoughtlet: “**SPREAD THIS MEME.**” With free flow of information, this will be the only value that remains—every other axiom will be cancelled out by its opposite, but the codon “don’t spread this meme” will, definitionally, not spread.”

~Hotel Concierge, The Tower

I was trying to make a memeplex that could compete in the, ahem, free marketplace of ideas, without a very firm understanding of what exactly that entailed, what the lay of the land was, what forces were at work in the background, or even, really, what it meant to compete in that environment. When I wrote the first draft of this essay, I wrote that

In many ways I was lucky, that I could have birthed something rather bad if it had taken longer to run off the rails and built up more inertia. But the truth was it never really had any inertia at all. I never came anywhere near constructing a functioning *bicycle*, much less *locomotive*.

This was, again, probably a good thing, given what I now know about memetics. That's not to say I don't still have an interest in pursuing the goals of the project, but the situation is a lot more complicated and difficult to get right than I had been thinking, and I'm being a lot more cautious moving forward. I want to make the world a better place, I don't want to delude myself into thinking I'm being helpful while producing something that makes the world worse.

This is going to be a constant theme throughout this set of however many long essays this ends up being, reality is more complicated and less easy to solve than it appears at first glance. I think in many ways this was a weakness to the framework generated by rationalist fiction, which presented many things as being simple puzzles which the characters could successfully solve and did so on the first try. Shikashi, actual reality is always more complicated than it first appears.

There are no easy hacks, the low hanging fruit has all been picked, there's no master set of techniques that if we just got everyone to adopt them, the world would be saved.

There is no consciousness expanding chemical solution or ideological superweapon that will save the world. There is no book or bible, no memeplex or manifesto, that contains the solution to humanity's ills. There is no easy answer.

All there is, is us.