Community and Cooperation

1. In Favor of Niceness, Community, and Civilization
2. Guided By The Beauty Of Our Weapons
3. The Ideology Is Not The Movement
4. Archipelago and Atomic Communitarianism
5. Meditations On Moloch
6. Five Planets In Search Of A Sci-Fi Story
7. It Was You Who Made My Blue Eyes Blue

In Favor of Niceness, Community, and Civilization

[Content warning: Discussion of social justice, discussion of violence, spoilers for Jacqueline Carey books.]

[Edit 10/25: This post was inspired by a debate with a friend of a friend on Facebook who has since become somewhat famous. I've renamed him here to "Andrew Cord" to protect his identity.]

I.

Andrew Cord criticizes me for my bold and controversial suggestion that maybe people should try to tell slightly fewer blatant hurtful lies:

> I just find it kind of darkly amusing and sad that the "rationalist community" loves "rationality is winning" so much as a tagline and yet are clearly not winning. And then complain about losing rather than changing their tactics to match those of people who are winning.
>
> Which is probably because if you *really* want to be the kind of person who wins you have to actually care about winning something, which means you have to have politics, which means you have to embrace "politics the mindkiller" and "politics is war and arguments are soldiers", and Scott would clearly rather spend the rest of his life losing than do this.
>
> That post [the one debunking false rape statistics] is exactly my problem with Scott. He seems to honestly think that it's a worthwhile use of his time, energy and mental effort to download evil people's evil worldviews into his mind and try to analytically debate them with statistics and cost-benefit analyses.
>
> He gets *mad* at people whom he detachedly intellectually agrees with but who are willing to back up their beliefs with war and fire rather than pussyfooting around with debate-team nonsense.
>
> It honestly makes me kind of sick. It is exactly the kind of thing that "social justice" activists like me *intend* to attack and "trigger" when we use "triggery" catchphrases about the mewling pusillanimity of privileged white allies.

In other words, if a fight is important to you, fight nasty. If that means lying, lie. If that means insults, insult. If that means silencing people, silence.

It always makes me happy when my ideological opponents come out and say eloquently and openly what I've always secretly suspected them of believing.

My natural instinct is to give some of the reasons why I think Andrew is wrong, starting with the history of the "noble lie" concept and moving on to some examples of why it didn't work very well, and why it might not be expected not to work so well in the future.

But in a way, that would be assuming the conclusion. I wouldn't be showing respect for Andrew's arguments. I wouldn't be going halfway to meet them on their own

terms.

The respectful way to rebut Andrew's argument would be to spread malicious lies about Andrew to a couple of media outlets, fan the flames, and wait for them to destroy his reputation. Then if the stress ends up bursting an aneurysm in his brain, I can dance on his grave, singing:

> ♪ ♬ I won this debate in a very effective manner. Now you can't argue in favor of nasty debate tactics any more ♬ ♪

I'm not going to do that, but if I *did* it's unclear to me how Andrew could object. I mean, he thinks that sexism is detrimental to society, so spreading lies and destroying people is justified in order to stop it. I think that discourse based on mud-slinging and falsehoods is detrimental to society. Therefore…

II.

But really, all this talk of lying and spreading rumors about people is – what was Andrew's terminology – "pussyfooting around with debate-team nonsense". You know who got things done? The IRA. They didn't agree with the British occupation of Northern Ireland and they weren't afraid to let people know in that very special way only a nail-bomb shoved through your window at night can.

Why not assassinate prominent racist and sexist politicians and intellectuals? I won't name names since that would be crossing a line, but I'm sure you can generate several of them who are sufficiently successful and charismatic that, if knocked off, there would not be an equally competent racist or sexist immediately available to replace them, and it would thus be a serious setback for the racism/sexism movement.

Other people can appeal to "the social contract" or "the general civilizational rule not to use violence", but not Andrew:

> I think that whether or not I use certain weapons has zero impact on whether or not those weapons are used against me, and people who think they do are either appealing to a kind of vague Kantian morality that I think is invalid or a specific kind of "honor among foes" that I think does not exist.

And don't give me that nonsense about the police. I'm sure a smart person like you can think of clever exciting new ways to commit the perfect murder. Unless you do not believe there will *ever* be an opportunity to defect unpunished, you need this sort of social contract to take you at least some of the way.

He continues:

> When Scott calls rhetorical tactics he dislikes "bullets" and denigrates them it actually hilariously plays right into this point…to be "pro-bullet" or "anti-bullet" is ridiculous. Bullets, as you say, are neutral. I am in favor of my side using bullets as best they can to destroy the enemy's ability to use bullets.
>
> In a war, a real war, a war for survival, you use all the weapons in your arsenal because you assume the enemy will use all the weapons in theirs. Because you understand that it IS a war.

There are a lot of things I am tempted to say to this.

Like “And that is why the United States immediately nukes every country it goes to war with.”

Or “And that is why the Geneva Convention was so obviously impossible that no one even bothered to attend the conference”.

Or “And that is why, to this very day, we solve every international disagreement through total war.”

Or “And that is why Martin Luther King was immediately reduced to a nonentity, and we remember the Weathermen as the sole people responsible for the success of the civil rights movement”

But I think what I am *actually* going to say is that, for the love of God, if you like bullets so much, stop using them as a metaphor for ‘spreading false statistics’ and go buy a gun.

III.

So let’s derive why violence is not in fact The One True Best Way To Solve All Our Problems. You can get most of this from Hobbes, but this blog post will be shorter.

Suppose I am a radical Catholic who believes all Protestants deserve to die, and therefore go around killing Protestants. So far, so good.

Unfortunately, there might be some radical Protestants around who believe all Catholics deserve to die. If there weren’t before, there probably are now. So they go around killing Catholics, we’re both unhappy and/or dead, our economy tanks, hundreds of innocent people end up as collateral damage, and our country goes down the toilet.

So we make an agreement: I won’t kill any more Catholics, you don’t kill any more Protestants. The specific Irish example was called the Good Friday Agreement and the general case is called “civilization”.

So then I try to destroy the hated Protestants using the government. I go around trying to pass laws banning Protestant worship and preventing people from condemning Catholicism.

Unfortunately, maybe the next government in power is a Protestant government, and they pass laws banning Catholic worship and preventing people from condemning Protestantism. No one can securely practice their own religion, no one can learn about other religions, people are constantly plotting civil war, academic freedom is severely curtailed, and once again the country goes down the toilet.

So again we make an agreement. I won’t use the apparatus of government against Protestantism, you don’t use the apparatus of government against Catholicism. The specific American example is the First Amendment and the general case is called “liberalism”, or to be dramatic about it, “civilization 2.0”

Every case in which both sides agree to lay down their weapons and be nice to each other has corresponded to spectacular gains by both sides and a new era of human flourishing.

“Wait a second, no!” someone yells. “I see where you’re going with this. You’re going to say that agreeing not to spread malicious lies about each other would also be a civilized and beneficial system. Like maybe the Protestants could stop saying that the Catholics worshipped the Devil, and the Catholics could stop saying the Protestants hate the Virgin Mary, and they could both relax the whole thing about the Jews baking the blood of Christian children into their matzah.

“But your two examples were about contracts written on paper and enforced by the government. So maybe a ‘no malicious lies’ amendment to the Constitution would work if it were enforceable, *which it isn’t*, but just *asking* people to stop spreading malicious lies is doomed from the start. The Jews will no doubt spread lies against *us*, so if we stop spreading lies about them, all we’re doing is abandoning an effective weapon against a religion I personally know to be heathenish! Rationalists should win, so put the blood libel on the front page of every newspaper!”

Or, as Andrew puts it:

> Whether or not I use certain weapons has zero impact on whether or not those weapons are used against me, and people who think they do are either appealing to a kind of vague Kantian morality that I think is invalid or a specific kind of “honor among foes” that I think does not exist.

So let’s talk about how beneficial game-theoretic equilibria can come to exist even in the absence of centralized enforcers. I know of two main ways: reciprocal communitarianism, and divine grace.

Reciprocal communitarianism is probably how altruism evolved. Some mammal started running TIT-FOR-TAT, the program where you cooperate with anyone whom you expect to cooperate with you. Gradually you form a successful community of cooperators. The defectors either join your community and agree to play by your rules or get outcompeted.

Divine grace is more complicated. I was tempted to call it “spontaneous order” until I remembered the rationalist proverb that if you don’t understand something, you need to call it by a term that reminds you that don’t understand it or else you’ll think you’ve explained it when you’ve just named it.

But consider the following: I am a pro-choice atheist. When I lived in Ireland, one of my friends was a pro-life Christian. I thought she was responsible for the unnecessary suffering of millions of women. She thought I was responsible for killing millions of babies. And yet she invited me over to her house for dinner without poisoning the food. And I ate it, and thanked her, and sent her a nice card, without smashing all her china.

Please try not to be insufficiently surprised by this. Every time a Republican and a Democrat break bread together with good will, it is a miracle. It is an equilibrium as beneficial as civilization or liberalism, which developed in the total absence of any central enforcing authority.

When you look for these equilibria, there are lots and lots. Andrew says there is no “honor among foes”, but if you read the *Iliad* or any other account of ancient warfare, there is practically nothing *but* honor among foes, and it wasn’t generated by some sort of Homeric version of the Geneva Convention, it just sort of happened. During World War I, the English and Germans spontaneously got out of their trenches and celebrated Christmas together with each other, and on the sidelines Andrew was

shouting “No! Stop celebrating Christmas! Quick, shoot them before they shoot you!” but they didn’t listen.

All I will say in way of explaining these miraculous equilibria is that they seem to have something to do with inheriting a cultural norm and not screwing it up. Punishing the occasional defector seems to be a big part of not screwing it up. How exactly that cultural norm came to be is less clear to me, but it might have something to do with the reasons why an entire civilization’s bureaucrats may suddenly turn 100% honest at the same time. I’m pretty sure I’m supposed to say the words timeless decision theory around this point too, and perhaps bring up the kind of Platonic contract that I have written about previously.

I think most of our useful social norms exist through a combination of divine grace and reciprocal communitarianism. To some degree they arise spontaneously and are preserved by the honor system. To another degree, they are stronger or weaker in different groups, and the groups that enforce them are so much more pleasant than the groups that don’t that people are willing to go along.

The norm against malicious lies follows this pattern. Politicians lie, but not *too much*. Take the top story on Politifact Fact Check today. Some Republican claimed his supposedly-maverick Democratic opponent actually voted with Obama’s economic policies 97 percent of the time. Fact Check explains that the statistic used was actually for *all* votes, not just economic votes, and that members of Congress typically have to have >90% agreement with their president because of the way partisan politics work. So it’s a lie, and is properly listed as one. But it’s a lie based on slightly misinterpreting a real statistic. He didn’t just totally make up a number. He didn’t even just make up something else, like “My opponent personally helped design most of Obama’s legislation”.

Even the guy in the fake rape statistics post lied less than he *possibly could have*. He got his fake numbers by conflating rapes per sex act with rapes per lifetime, and it’s really hard for me to imagine someone doing that by anything resembling accident. But he couldn’t bring himself to go the extra step and just totally make up numbers with no grounding whatsoever. And part of me wonders: why not? If you’re going to use numbers you know are false to destroy people, why is it better to derive the numbers through a formula you know is incorrect, than to just skip the math and make the numbers up in the first place? “The FBI has determined that no false rape claims have ever been submitted, my source is an obscure report they published, when your local library doesn’t have it you will just accept that libraries can’t have all books, and suspect nothing.”

This would have been a *more believable* claim than the one he made. Because he showed his work, it was easy for me to debunk it. If he had just said it was in some obscure report, I wouldn’t have gone through the trouble. So why did he go the harder route?

People *know* lying is wrong. They know if they lied they would be punished. More ~~spontaneous social order~~ miraculous divine grace. And so they want to hedge their bets, be able to say “Well, I didn’t exactly *lie*, per se.”

And this is good! We *want* to make it politically unacceptable to have people say that Jews bake the blood of Christian children into their matzah. Now we build on that success. We start hounding around the edges of currently acceptable lies. “Okay, you didn’t *literally* make up your statistics, but you still lied, and you still should be cast

out from the community of people who have reasonable discussions and never trusted by anyone again."

It might not totally succeed in making a new norm against this kind of thing. But at least it will prevent other people from seeing their success, taking heart, and having the number of lies which are socially acceptable gradually *advance*.

So much for protecting what we have been given by divine grace. But there is also reciprocal communitarianism to think of.

I seek out people who signal that they want to discuss things honestly and rationally. Then I try to discuss things honestly and rationally with those people. I try to concentrate as much of my social interaction there as possible.

So far this project is going pretty well. My friends are nice, my romantic relationships are low-drama, my debates are productive and I am learning so, so much.

And people think "Hm, I could hang out at 4Chan and be called a 'fag'. Or I could hang out at Slate Star Codex and discuss things rationally and learn a lot. And if I want to be allowed in, all I have to do is not be an intellectually dishonest jerk."

And so our community grows. And all over the world, the mysterious divine forces favoring honest and kind equilibria gain a little bit more power over the mysterious divine forces favoring lying and malicious equilibria.

Andrew thinks I am trying to fight all the evils of the world, and doing so in a stupid way. But sometimes I just want to cultivate my garden.

IV.

Andrew goes on to complain:

> Scott…seems to [dispassionately debate] evil people's evil worldviews …with statistics and cost-benefit analyses.
>
> He gets *mad* at people whom he detachedly intellectually agrees with but who are willing to back up their beliefs with war and fire rather than pussyfooting around with debate-team nonsense.

I accept this criticism as an accurate description of what I do.

Compare to the following two critiques: "The Catholic Church wastes so much energy getting upset about heretics who believe *mostly* the same things as they do, when there are literally *millions* of Hindus over in India who don't believe in Catholicism *at all*! What dumb priorities!"

Or "How could Joseph McCarthy get angry about a couple of people who *might* have been Communists in the US movie industry, when over in Moscow there were *thousands* of people who were openly *super* Communist *all the time*?"

There might be foot-long giant centipedes in the Amazon, but I am a lot more worried about boll weevils in my walled garden.

Creationists lie. Homeopaths lie. Anti-vaxxers lie. This is part of the Great Circle of Life. It is not necessary to call out every lie by a creationist, because the sort of person who is still listening to creationists is not the sort of person who is likely to be

moved by call-outs. There is a role for *organized* action against creationists, like preventing them from getting their opinions taught in schools, but the marginal blog post "debunking" a creationist on something is a waste of time. Everybody who wants to discuss things rationally has already formed a walled garden and locked the creationists outside of it.

Anti-Semites fight nasty. The Ku Klux Klan fights nasty. Neo-Nazis fight nasty. We dismiss them with equanimity, in accordance with the ancient proverb: "Haters gonna hate". There is a role for *organized* opposition to these groups, like making sure they can't actually terrorize anyone, but the marginal blog post condemning Nazism is a waste of time. Everybody who wants to discuss things charitably and compassionately has already formed a walled garden and locked the Nazis outside of it.

People who want to discuss things rationally and charitably have not yet looked over the false rape statistics article and decided to lock Charles Clymer out of their walled garden.

He is not a heathen, he is a heretic. He is not a foreigner, he is a traitor. He comes in talking all liberalism and statistics, and then he betrays the signals he has just sent. He is not just some guy who defects in the Prisoner's Dilemma. He is the guy who defects while wearing the "I COOPERATE IN PRISONERS DILEMMAS" t-shirt.

What really, *really* bothered me wasn't Clymer at all: it was that *rationalists* were taking him seriously. Smart people, kind people! I even said so in my article. Boll weevils in our beautiful walled garden!

Why am I always harping on feminism? I feel like we've got a good thing going, we've ratified our Platonic contract to be intellectually honest and charitable to each other, we are going about perma-cooperating in the Prisoner's Dilemma and reaping gains from trade.

And then someone says "Except that of course regardless of all that I reserve the right to still use lies and insults and harassment and dark epistemology to spread feminism". Sometimes they do this explicitly, like Andrew did. Other times they use a more nuanced argument like "Surely you didn't think the same rules against lies and insults and harassment should apply to oppressed and privileged people, did you?" And other times they don't say anything, but just show their true colors by reblogging an awful article with false statistics.

(and still other times they don't do any of this and they are wonderful people whom I am glad to know)

But then someone else says "Well, if they get their exception, I deserve my exception," and then someone else says "Well, if those two get exceptions, I'm out", and *you have no idea how difficult it is to successfully renegotiate the terms of a timeless Platonic contract that doesn't literally exist*.

No! I am Exception Nazi! NO EXCEPTION FOR YOU! Civilization didn't conquer the world by forbidding you to murder your enemies *unless* they are actually unrighteous in which case go ahead and kill them all. Liberals didn't give their lives in the battle against tyranny to end discrimination against all religions *except* Jansenism because seriously fuck Jansenists. Here we have built our Schelling fence and here we are defending it to the bitter end.

V.

Contrary to how it may appear, I am not trying to doom feminism.

Feminists like to mock the naivete of anyone who says that classical liberalism would suffice to satisfy feminist demands. And true, you cannot simply assume Adam Smith and derive Andrea Dworkin. Not being an asshole to women and not writing laws declaring them officially inferior are both good starts, but it not enough if there's still cultural baggage and entrenched gender norms.

But here I am, defending this principle – kind of a supercharged version of liberalism – of "It is not okay to use lies, insults, and harassment against people, even if it would help you enforce your preferred social norms."

And I notice that this gets us a heck of a lot closer to feminism than Andrew's principle of "Go ahead and use lies, insults, and harassment if they are effective ways to enforce your preferred social norms."

Feminists are very concerned about slut-shaming, where people harass women who have too much premarital sex. They point out that this is very hurtful to women, that men might underestimate the amount of hurt it causes women, and that the standard-classical-liberal solution of removing relevant government oppression does nothing. All excellent points.

But one assumes the harassers think that women having premarital sex is detrimental to society. So they apply their general principle: "I should use lies, insults, and harassment to enforce my preferred social norms."

But this is the principle Andrew is asserting, against myself and liberalism.

Feminists think that women should be free from fear of rape, and that, if raped, no one should be able to excuse themselves with "well, she was asking for it".

But this is the same anti-violence principle as saying that the IRA shouldn't throw nail-bombs through people's windows or that, nail bombs having been thrown, the IRA can't use as an excuse "Yeah, well, they were complicit with the evil British occupation, they deserved it." Again, I feel like I'm defending this principle a whole lot more strongly and consistently than Andrew is.

Feminists are, shall we say, divided about transgender people, but let's allow that the correct solution is to respect their rights.

When I was young and stupid, I used to believe that transgender was really, really dumb. That they were looking for attention or making it up or something along those lines.

Luckily, since I was a classical liberal, my reaction to this mistake was – to not bother them, and to get very very angry at people who did bother them. I got upset with people trying to fire Phil Robertson for being homophobic even though homophobia is stupid. You better bet I also got upset with people trying to fire transgender people back when I thought transgender was stupid.

And then I grew older and wiser and learned – hey, transgender isn't stupid at all, they have very important reasons for what they do and go through and I was atrociously wrong. And I said a mea culpa.

But it could have been worse. I didn't like transgender people, and so I *left them alone while still standing up for their rights*. My epistemic structure *failed gracefully*. For anyone who's not overconfident, and so who expects massive epistemic failure on a variety of important issues all the time, graceful failure modes are a *really important feature* for an epistemic structure to have.

God only knows what Andrew would have done, if through bad luck he had accidentally gotten it into his head that transgender people are bad. From his own words, we know he wouldn't be "pussyfooting around with debate-team nonsense".

I admit there are many feminist principles that cannot be derived from, or are even opposed to my own liberal principles. For example, some feminists have suggested that pornography be banned because it increases the likelihood of violence against women. Others suggest that research into gender differences should be banned, or at least we should stigmatize and harass the researchers, because any discoveries made might lend aid and comfort to sexists.

To the first, I would point out that there is now strong evidence that pornography, especially violent objectifying pornography, very significantly decreases violence against women. I would ask them whether they're happy that we did the nice liberal thing and waited until all the evidence came in so we could discuss it rationally, rather than immediately moving to harass and silence anyone taking the pro-pornography side.

And to the second, well, we have a genuine disagreement. But I wonder whether they would prefer to discuss that disagreement reasonably, or whether we should both try to harass and destroy the other until one or both of us are too damaged to continue the struggle.

And if feminists agree to have that reasonable discussion, but lose, I would tell them that they get a consolation prize. Having joined liberal society, they can be sure that no matter what those researchers find, I and all of their new liberal-society buddies will fight tooth and nail against anyone who uses any tiny differences those researchers find to challenge the central liberal belief that everyone of every gender has basic human dignity. Any victory for me is going to be a victory for feminists as well; maybe not a perfect victory, but a heck of a lot better than what they have right now.

VI.

I am not trying to fight all the evils of the world. I am just trying to cultivate my garden.

And you argue: "But isn't that selfish and oppressive and privileged? Isn't that confining everyone outside of your walled garden to racism and sexism and nastiness?

But there is a famous comic which demonstrates what can happen to certain walled gardens.

Why yes, it does sound like I'm making the unshakeable assumption that liberalism always wins, doesn't it? That people who voluntarily relinquish certain forms of barbarism will be able to gradually expand their territory against the hordes outside, instead of immediately being conquered by their less scrupulous neighbors? And it looks like Andrew isn't going to let that assumption pass.

He writes:

> The *whole history* of why the institutional Left in our society is a party of toothless, spineless, gutless losers and they've spent two generations doing nothing but lose.

One is reminded of the old joke about the Nazi papers. The rabbi catches an old Jewish man reading the Nazi newspaper and demands to know how he could look at such garbage. The man answers "When I read our Jewish newpapers, the news is so depressing – oppression, death, genocide! But here, everything is great! We control the banks, we control the media. Why, just yesterday they said we had a plan to kick the Gentiles out of Germany entirely!"

And I have two thoughts about this.

First, it argues that "Evil people are doing evil things, so we are justified in using any weapons we want to stop them, no matter how nasty" suffers from a certain flaw. Everyone believes their enemies are evil people doing evil things. If you're a Nazi, you are just defending yourself, in a very proportionate manner, against the Vast Jewish Conspiracy To Destroy All Germans.

But second, before taking Andrew's words for how disastrously liberalism is doing, we should check the newspapers put out by liberalism's enemies. Here's Mencius Moldbug:

> Cthulhu may swim slowly. But he only swims left. Isn't that interesting?
>
> In each of the following conflicts in Anglo-American history, you see a victory of left over right: the English Civil War, the so-called "Glorious Revolution," the American Revolution, the American Civil War, World War I, and World War II. Clearly, if you want to be on the winning team, you want to start on the left side of the field.
>
> Where is the John Birch Society, now? What about the NAACP? Cthulhu swims left, and left, and left. There are a few brief periods of true reaction in American history – the post-Reconstruction era or Redemption, the Return to Normalcy of Harding, and a couple of others. But they are unusual and feeble compared to the great leftward shift. McCarthyism is especially noticeable as such. And you'll note that McCarthy didn't exactly win.
>
> In the history of American democracy, if you take the mainstream political position (Overton Window, if you care) at time T1, and place it on the map at a later time T2, T1 is always way to the right, near the fringe or outside it. So, for instance, if you take the average segregationist voter of 1963 and let him vote in the 2008 election, he will be way out on the wacky right wing. Cthulhu has passed him by.

I've got to say Mencius makes a much more convincing argument than Andrew does.

Robert Frost says "A liberal is a man too broad-minded to take his own side in a quarrel". Ha ha ha.

And yet, outside of Saudi Arabia you'll have a hard time finding a country that doesn't at least pay lip service to liberal ideas. Stranger still, many of those then go on to *actually implement them*, either voluntarily or after succumbing to strange pressures

they don't understand. In particular, the history of the past few hundred years in the United States has been a history of decreasing censorship and increasing tolerance.

Contra the Reactionaries, feminism isn't an exception to that, it's a casualty of it. 1970s feminists were saying that all women need to rise up and smash the patriarchy, possibly with literal smashing-implements. 2010s feminists are saying that if some women want to be housewives, that's great and their own choice because in a liberal society everyone should be free to pursue their own self-actualization.

And that has *corresponded to* spectacular successes of the specific causes liberals like to push, like feminism, civil rights, gay marriage, et cetera, et cetera, et cetera.

A liberal is a man too broad-minded to take his own side in a quarrel. And yet when liberals enter quarrels, they always win. Isn't that interesting?

VII.

Andrew thinks that liberals who voluntarily relinquish any form of fighting back are just ignoring perfectly effective weapons. I'll provide the quote:

> In a war, a real war, a war for survival, you use all the weapons in your arsenal because you assume the enemy will use all the weapons in theirs. Because you understand that it IS a war… Any energy spent mentally debating how, in a perfect world run by a Lawful Neutral Cosmic Arbiter that will never exist, we could settle wars without bullets is energy you could better spend down at the range improving your marksmanship… I am amazed that the "rationalist community" finds it to still be so opaque.

Let me name some other people who mysteriously managed to miss this perfectly obvious point.

The early Christian Church had the slogan "resist not evil" (Matthew 5:39), and indeed, their idea of Burning The Fucking System To The Ground was to go unprotestingly to martyrdom while publicly forgiving their executioners. They were up against the Roman Empire, possibly the most effective military machine in history, ruled by some of the cruelest men who have ever lived. By Andrew's reckoning, this should have been the biggest smackdown in the entire history of smackdowns.

And it kind of was. Just not the way most people expected.

Mahatma Gandhi said "Non-violence is the greatest force at the disposal of mankind. It is mightier than the mightiest weapon of destruction devised by the ingenuity of man." Another guy who fought one of the largest empires ever to exist and won resoundingly. And he was pretty insistent on truth too: "Non-violence and truth are inseparable and presuppose one another."

Also skilled at missing the obvious: Martin Luther King. Desmond Tutu. Aung San Suu Kyi. Nelson Mandela was smart and effective at the beginning of his career, but fell into a pattern of missing the obvious when he was older. Maybe it was Alzheimers.

Of course, there are counterexamples. Jews who nonviolently resisted the Nazis didn't have a very good track record. You need a certain pre-existing level of civilization for liberalism to be a good idea, and a certain pre-existing level of liberalism for supercharged liberalism where you don't spread malicious lies and harass other

people to be a good idea. You need to have pre-existing community norms in place before trying to summon mysterious beneficial equilibria.

So perhaps I am being too harsh on Andrew, to contrast him with Aung San Suu Kyi and her ilk. After all, all Aung San Suu Kyi had to do was fight the Burmese junta, a cabal of incredibly brutal military dictators who killed several thousand people, tortured anyone who protested against them, and sent eight hundred thousand people they just didn't like to forced labor camps. Andrew has to deal with *people on Facebook who aren't as feminist as he is*. Clearly this requires much stronger measures!

VIII.

Liberalism does not conquer by fire and sword. Liberalism conquers by communities of people who agree to play by the rules, slowly growing until eventually an equilibrium is disturbed. Its battle cry is not "Death to the unbelievers!" but "If you're nice, you can join our cuddle pile!"

But some people, through lack of imagination, fail to find this battle cry sufficiently fear-inspiring.

I hate to invoke fictional evidence, especially since perhaps Andrew's strongest point is that the real world doesn't work like fiction. But these people need to read Jacqueline Carey's *Kushiel's Avatar*.

Elua is the god of kindness and flowers and free love. All the other gods are gods of blood and fire, and Elua is just like "Love as thou wilt" and "All knowlege is worth having". He is the patron deity of exactly the kind of sickeningly sweet namby-pamby charitable liberalism that Andrew is complaining about.

And there is a certain commonality to a lot of the Kushiel books, where some tyrant or sorcerer thinks that a god of flowers and free love will be a pushover, and starts harassing his followers. And the only Eluite who shows up to stop him is Phèdre nó Delaunay, and the tyrant thinks "Ha! A woman, who doesn't even know how to fight, doesn't have any magic! What a wuss!"

But here is an important rule about dealing with fantasy book characters.

If you ever piss off Sauron, you should probably find the Ring of Power and take it to Mount Doom.

If you ever get piss off Voldemort, you should probably start looking for Horcruxes.

If you ever piss off Phèdre nó Delaunay, *run and never stop running*.

Elua is the god of flowers and free love and he is terrifying. If you oppose him, there will not be enough left of you to bury, and it will not matter because there will not be enough left of your city to bury you in.

And Jacqueline Carey and Mencius Moldbug are both wiser than Andrew Cord.

Carey portrays liberalism as Elua, a terrifying unspeakable Elder God who is fundamentally good.

Moldbug portrays liberalism as Cthulhu, a terrifying unspeakable Elder God who is fundamentally evil.

But Andrew? He *doesn't even seem to realize liberalism is a terrifying unspeakable Elder God at all*. It's like, *what?*

Andrew is the poor shmuck who is sitting there saying "Ha ha, a god who doesn't even control any hell-monsters or command his worshippers to become killing machines. What a weakling! This is going to be so easy!"

And you want to scream: "THERE IS ONLY ONE WAY THIS CAN POSSIBLY END AND IT INVOLVES YOU BEING EATEN BY YOUR OWN LEGIONS OF DEMONAICALLY CONTROLLED ANTS"

(uh, spoilers)

Guided By The Beauty Of Our Weapons

[Content note: kind of talking around Trump supporters and similar groups as if they're not there.]

I.

Tim Harford writes The Problem With Facts, which uses Brexit and Trump as jumping-off points to argue that people are mostly impervious to facts and resistant to logic:

> All this adds up to a depressing picture for those of us who aren't ready to live in a post-truth world. Facts, it seems, are toothless. Trying to refute a bold, memorable lie with a fiddly set of facts can often serve to reinforce the myth. Important truths are often stale and dull, and it is easy to manufacture new, more engaging claims. And giving people more facts can backfire, as those facts provoke a defensive reaction in someone who badly wants to stick to their existing world view. "This is dark stuff," says Reifler. "We're in a pretty scary and dark time."

He admits he has no easy answers, but cites some studies showing that "scientific curiosity" seems to help people become interested in facts again. He thinks maybe we can inspire scientific curiosity by linking scientific truths to human interest stories, by weaving compelling narratives, and by finding "a Carl Sagan or David Attenborough of social science".

I think this is generally a good article and makes important points, but there are three issues I want to highlight as possibly pointing to a deeper pattern.

First, the article makes the very strong claim that "facts are toothless" – then tries to convince its readers of this using facts. For example, the article highlights a study by Nyhan & Reifler which finds a "backfire effect" – correcting people's misconceptions only makes them cling to those misconceptions more strongly. Harford expects us to be impressed by this study. But how is this different from all of those social science facts to which he believes humans are mostly impervious?

Second, Nyhan & Reifler's work on the backfire effect is probably not true. The original study establishing its existence failed to replicate (see eg Porter & Wood, 2016). This isn't directly contrary to Harford's argument, because Harford doesn't cite the original study – he cites a slight extension of it done a year later by the same team that comes to a slightly different conclusion. But given that the entire field is now in serious doubt, I feel like it would have been judicious to mention some of this in the article. This is especially true given that the article itself is about the way that false ideas spread by people never double-checking their beliefs. It seems to me that if you believe in an epidemic of falsehood so widespread that the very ability to separate fact from fiction is under threat, it ought to inspire a state of CONSTANT VIGILANCE, where you obsessively question each of your beliefs. Yet Harford writes an entire article about a worldwide plague of false beliefs without mustering enough vigilance to see if the relevant studies are true or not.

Third, Harford describes his article as being about *agnotology*, "the study of how ignorance is deliberately produced". His key example is tobacco companies sowing doubt about the negative health effects of smoking – for example, he talks about tobacco companies sponsoring (accurate) research into all of the non-smoking-related

causes of disease so that everyone focused on those instead. But his solution – telling engaging stories, adding a human interest element, enjoyable documentaries in the style of Carl Sagan – seems unusually unsuited to the problem. The National Institute of Health can make an engaging human interest documentary about a smoker who got lung cancer. And the tobacco companies can make an engaging human interest documentary about a guy who got cancer because of asbestos, then was saved by tobacco-sponsored research. Opponents of Brexit can make an engaging documentary about all the reasons Brexit would be bad, and then proponents of Brexit can make an engaging documentary about all the reasons Brexit would be good. If you get good documentary-makers, I assume both will be equally convincing regardless of what the true facts are.

All three of these points are slightly unfair. The first because Harford's stronger statements about facts are probably exaggerations, and he just meant that in *certain cases* people ignore evidence. The second because the specific study cited wasn't the one that failed to replicate and Harford's thesis might be that it was different enough from the original that it's probably true. And the third because the documentaries were just one idea meant to serve a broader goal of increasing "scientific curiosity", a construct which has been shown in studies to be helpful in getting people to believe true things.

But I worry that taken together, they suggest an unspoken premise of the piece. It isn't that *people* are impervious to facts. Harford doesn't expect his reader to be impervious to facts, he doesn't expect documentary-makers to be impervious to facts, and he certainly doesn't expect *himself* to be impervious to facts. The problem is that there's some weird tribe of fact-immune troglodytes out there, going around refusing vaccines and voting for Brexit, and the rest of us have to figure out what to do about them. The fundamental problem is one of *transmission*: how can we make knowledge percolate down from the fact-loving elite to the fact-impervious masses?

And I don't want to condemn this too hard, because it's obviously true up to a point. Medical researchers have lots of useful facts about vaccines. Statisticians know some great facts about the link between tobacco and cancer (shame about Ronald Fisher, though). Probably there are even some social scientists who have a fact or two.

Yet as I've argued before, excessive focus on things like vaccine denialists teaches the wrong habits. It's a desire to take a degenerate case, the rare situation where one side is obviously right and the other bizarrely wrong, and make it into the flagship example for modeling all human disagreement. Imagine a theory of jurisprudence designed only to smack down sovereign citizens, or a government pro-innovation policy based entirely on warning inventors against perpetual motion machines.

And in this wider context, part of me wonders if the focus on transmission is part of the problem. Everyone from statisticians to Brexiteers knows that they are right. The only remaining problem is how to convince others. Go on Facebook and you will find a million people with a million different opinions, each confident in her own judgment, each zealously devoted to informing everyone else.

Imagine a classroom where everyone believes they're the teacher and everyone else is students. They all fight each other for space at the blackboard, give lectures that nobody listens to, assign homework that nobody does. When everyone gets abysmal test scores, one of the teachers has an idea: *I need a more engaging curriculum*. Sure. That'll help.

II.

A new Nathan Robinson article: Debate Vs. Persuasion. It goes through the same steps as the Harford article, this time from the perspective of the political Left. Deploying what Robinson calls "Purely Logical Debate" against Trump supporters hasn't worked. Some leftists think the answer is violence. But this may be premature; instead, we should try the tools of rhetoric, emotional appeal, and other forms of discourse that aren't Purely Logical Debate. In conclusion, Bernie Would Have Won.

> I think giving up on argumentation, reason, and language, just because Purely Logical Debate doesn't work, is a mistake. It's easy to think that if we can't convince the right with facts, there's no hope at all for public discourse. But this might not suggest anything about the possibilities of persuasion and dialogue. Instead, it might suggest that mere facts are rhetorically insufficient to get people excited about your political program.

The resemblance to Harford is obvious. You can't convince people with facts. But you *might* be able to convince people with facts carefully intermixed with human interest, compelling narrative, and emotional appeal.

Once again, I think this is generally a good article and makes important points. But I still want to challenge whether things are quite as bad as it says.

Google "debating Trump supporters is", and you realize where the article is coming from. It's page after page of "debating Trump supporters is pointless", "debating Trump supporters is a waste of time", and "debating Trump supporters is like [funny metaphor for thing that doesn't work]". The overall picture you get is of a world full of Trump opponents and supporters debating on every street corner, until finally, after months of banging their heads against the wall, everyone collectively decided it was futile.

Yet I have the opposite impression. Somehow a sharply polarized country went through a historically divisive election with *essentially no debate taking place*.

Am I about to No True Scotsman the hell out of the word "debate"? Maybe. But I feel like in using the exaggerated phrase "Purely Logical Debate, Robinson has given me leave to define the term as strictly as I like. So here's what I think are minimum standards to deserve the capital letters:

1. Debate where two people with opposing views are *talking* to each other (or writing, or IMing, or some form of bilateral communication). Not a pundit putting an article on *Huffington Post* and demanding Trump supporters read it. Not even a Trump supporter who comments on the article with a counterargument that the author will never read. Two people who have chosen to engage and to listen to one another.

2. Debate where both people want to be there, and have chosen to enter into the debate in the hopes of getting something productive out of it. So not something where someone posts a "HILLARY IS A CROOK" meme on Facebook, someone gets really angry and lists all the reasons Trump is an even bigger crook, and then the original poster gets angry and has to tell them why they're wrong. Two people who have made it their business to come together at a certain time in order to compare opinions.

3. Debate conducted in the spirit of mutual respect and collaborative truth-seeking. Both people reject personal attacks or 'gotcha' style digs. Both people understand that the other person is *around* the same level of intelligence as they are and may

have some useful things to say. Both people understand that they themselves might have some false beliefs that the other person will be able to correct for them. Both people go into the debate with the hope of convincing their opponent, but not completely rejecting the possibility that their opponent might convince them also.

4. Debate conducted outside of a high-pressure point-scoring environment. No audience cheering on both participants to respond as quickly and bitingly as possible. If it can't be done online, at least do it with a smartphone around so you can open Wikipedia to resolve simple matters of fact.

5. Debate where both people agree on what's being debated and try to stick to the subject at hand. None of this "I'm going to vote Trump because I think Clinton is corrupt" followed by "Yeah, but Reagan was even worse and that just proves you Republicans are hypocrites" followed by "*We're* hypocrites? You Democrats claim to support women's rights but you love Muslims who make women wear headscarves!" Whether or not it's hypocritical to "support women's rights" but "love Muslims", it doesn't seem like anyone is even *trying* to change each other's mind about Clinton at this point.

These to me seem like the *bare minimum* conditions for a debate that could possibly be productive.

(and while I'm asking for a pony on a silver platter, how about both people have to read *How To Actually Change Your Mind* first?)

Meanwhile, in reality…

If you search "debating Trump supporters" without the "is", your first result is this video, where some people with a microphone corner some other people at what looks like a rally. I can't really follow the conversation because they're all shouting at the same time, but I can make out somebody saying 'Republicans give more to charity!' and someone else responding 'That's cause they don't do anything at their jobs!'". Okay.

The second link is this podcast where a guy talks about debating Trump supporters. After the usual preface about how stupid they were, he describes a typical exchange – "It's kind of amazing how they want to go back to the good old days…Well, when I start asking them 'You mean the good old days when 30% of the population were in unions'…they never seem to like to hear that!…so all this unfettered free market capitalism has got to go bye-bye. They don't find comfort in that idea either. It's amazing. I can say I now know what cognitive dissonance feels like on someone's face." I'm glad time travel seems to be impossible, because otherwise I would be tempted to warp back and change my vote to Trump just to spite this person.

The third link is Vanity Fair's "Foolproof Guide To Arguing With Trump Supporters", which suggests "using their patriotism against them" by telling them that wanting to "curtail the rights and privileges of certain of our citizens" is un-American.

I worry that people do this kind of thing every so often. Then, when it fails, they conclude "Trump supporters are immune to logic". This is much like observing that Republicans go out in the rain without melting, and concluding "Trump supporters are immortal".

Am I saying that if you met with a conservative friend for an hour in a quiet cafe to talk over your disagreements, they'd come away convinced? No. I've changed my

mind on various things during my life, and it was never a single moment that did it. It was more of a series of different things, each taking me a fraction of the way. As the old saying goes, “First they ignore you, then they laugh at you, then they fight you, then they fight you half-heartedly, then they’re neutral, then they then they grudgingly say you might have a point even though you’re annoying, then they say on balance you’re mostly right although you ignore some of the most important facets of the issue, then you win.”

There might be a parallel here with the one place I see something like Purely Logical Debate on a routine basis: cognitive psychotherapy. I know this comparison sounds crazy, because psychotherapy is supposed to be the opposite of a debate, and trying to argue someone out of their delusions or depression inevitably fails. The rookiest of all rookie therapist mistakes is to say “FACT CHECK: The patient says she is a loser who everybody hates. PsychiaFact rates this claim: PANTS ON FIRE.”

But in other ways it’s a lot like the five points above. You have two people who disagree – the patient thinks she’s a worthless loser who everyone hates, and the therapist thinks maybe not. They meet together in a spirit of voluntary mutual inquiry, guaranteed safe from personal attacks like “You’re crazy!”. Both sides go over the evidence together, sometimes even agreeing on explicit experiments like “Ask your boyfriend tonight whether he hates you or not, predict beforehand what you think he’s going to say, and see if your prediction is accurate”. And both sides approach the whole process suspecting that they’re right but admitting the possibility that they’re wrong (very occasionally, after weeks of therapy, I realize that frick, everyone really *does* hate my patient. Then we switch strategies to helping her with social skills, or helping her find better friends).

And contrary to what you see in movies, this doesn’t usually give a single moment of blinding revelation. If you spent your entire life talking yourself into the belief that you’re a loser and everyone hates you, no single fact or person is going to talk you out of it. But after however many months of intensive therapy, sometimes someone who was *sure* that they were a loser is now *sort of questioning* whether they’re a loser, and has the mental toolbox to take things the rest of the way themselves.

This was also the response I got when I tried to make an anti-Trump case on this blog. I don’t think there were any sudden conversions, but here were some of the positive comments I got from Trump supporters:

— “This is a compelling case, but I’m still torn.”

— “This contains the most convincing arguments for a Clinton presidency I have ever seen. But, perhaps also unsurprisingly, while it did manage to shift some of my views, it did not succeed in convincing me to change my bottom line.”

— “This article is perhaps the best argument I have seen yet for Hillary. I found myself nodding along with many of the arguments, after this morning swearing that there was nothing that could make me consider voting for Hillary…the problem in the end was that it wasn’t enough.”

— “The first coherent article I’ve read justifying voting for Clinton. I don’t agree with your analysis of the dollar “value” of a vote, but other than that, something to think about.”

— “Well I don’t like Clinton at all, and I found this essay reasonable enough. The argument from continuity is probably the best one for voting Clinton if you don’t

particularly love any of her policies or her as a person. Trump is a wild card, I must admit."

— As an orthodox Catholic, you would probably classify me as part of your conservative audience…I certainly concur with both the variance arguments and that he's not conservative by policy, life, or temperament, and I will remain open to hearing what you have to say on the topic through November.

— "I've only come around to the 'hold your nose and vote Trump' camp the past month or so…I won't say [you] didn't make me squirm, but I'm holding fast to my decision."

These are the people you say are completely impervious to logic so don't even try? It seems to me like this argument was one of not-so-many straws that might have broken some camels' backs if they'd been allowed to accumulate. And the weird thing is, when I re-read the essay I notice a lot of flaws and things I wish I'd said differently. I don't think it was an exceptionally good argument. I think it was…an argument. It was something more than saying "You think the old days were so great, but the old days had labor unions, CHECKMATE ATHEISTS". This isn't what you get when you do a splendid virtuouso perfomance. This is what you get *when you show up*.

(and lest I end up 'objectifying' Trump supporters as prizes to be won, I'll add that in the comments some people made pro-Trump arguments, and two people who were previously leaning Clinton said that they were feeling uncomfortably close to being convinced)

Another SSC story. I keep trying to keep "culture war"-style political arguments from overrunning the blog and subreddit, and every time I add restrictions a bunch of people complain that this is the only place they can go for that. Think about this for a second. A heavily polarized country of three hundred million people, split pretty evenly into two sides and obsessed with politics, blessed with the strongest free speech laws in the world, and people are complaining that I can't change my comment policy because this one small blog is *the only place they know where they can debate people from the other side*.

Given all of this, I reject the argument that Purely Logical Debate has been tried and found wanting. Like GK Chesterton, I think it has been found difficult and left untried.

III.

Therapy might change minds, and so might friendly debate among equals, but neither of them scales very well. Is there anything that big fish in the media can do beyond the transmission they're already trying?

Let's go back to that Nyhan & Reifler study which found that fact-checking backfired. As I mentioned above, a replication attempt by Porter & Wood found the opposite. This could have been the setup for a nasty conflict, with both groups trying to convince academia and the public that they were right, or even accusing the other of scientific malpractice.

Instead, something great happened. All four researchers decided to work together on an "adversarial collaboration" – a bigger, better study where they all had input into the methodology and they all checked the results independently. The collaboration found that fact-checking generally didn't backfire in most cases. All four of them used

their scientific clout to publicize the new result and launch further investigations into the role of different contexts and situations.

Instead of treating disagreement as demonstrating a need to transmit their own opinion more effectively, they viewed it as demonstrating a need to collaborate to investigate the question together.

And yeah, part of it was that they were all decent scientists who respected each other. But they didn't *have* to be. If one team had been total morons, and the other team was secretly laughing at them the whole time, the collaboration still would have worked. All required was an assumption of good faith.

A while ago I blogged about a journalistic spat between German Lopez and Robert VerBruggen on gun control. Lopez wrote a voxsplainer citing some statistics about guns. VerBruggen wrote a piece at National Review saying that some of the statistics were flawed. German fired back (pun not intended) with an article claiming that VerBruggen was ignoring better studies.

(Then I yelled at both of them, as usual.)

Overall the exchange was in the top 1% of online social science journalism – by which I mean it included at least one statistic and at some point that statistic was superficially examined. But in the end, it was still just two people arguing with one another, each trying to transmit his superior knowledge to each other and the reading public. As good as it was, it didn't meet my five standards above – and nobody expected it to.

But now I'm thinking – what would have happened if Lopez and VerBruggen had joined together in an adversarial collaboration? Agreed to work together to write an article on gun statistics, with nothing going into the article unless they both approved, and then they both published that article on their respective sites?

This seems like a mass media equivalent of shifting from Twitter spats to serious debate, from transmission mindset to collaborative truth-seeking mindset. The adversarial collaboration model is just the first one to come to mind right now. I've blogged about others before – for example, bets, prediction markets, and calibration training.

The media already spends a lot of effort *recommending* good behavior. What if they tried *modeling* it?

IV.

The bigger question hanging over all of this: "Do we *have* to?"

Harford's solution – compelling narratives and documentaries – sounds easy and fun. Robinson's solution – rhetoric and emotional appeals – also sounds easy and fun. Even the solution Robinson rejects – violence – is easy, and fun for a certain type of person. All three work on pretty much anybody.

Purely Logical Debate is difficult and annoying. It doesn't scale. It only works on the subset of people who are willing to talk to you in good faith and smart enough to understand the issues involved. And even then, it only works glacially slowly, and you win only partial victories. What's the point?

Logical debate has one advantage over narrative, rhetoric, and violence: it's an *asymmetric weapon*. That is, it's a weapon which is stronger in the hands of the good guys than in the hands of the bad guys. In ideal conditions (which may or may not ever happen in real life) – the kind of conditions where everyone is charitable and intelligent and wise – the good guys will be able to present stronger evidence, cite more experts, and invoke more compelling moral principles. The whole point of logic is that, when done right, it can only prove things that are true.

Violence is a *symmetric weapon*; the bad guys' punches hit just as hard as the good guys' do. It's true that hopefully the good guys will be more popular than the bad guys, and so able to gather more soldiers. But this doesn't mean violence itself is asymmetric – the good guys will only be more popular than the bad guys insofar as their ideas have previously spread through some means other than violence. Right now antifascists outnumber fascists and so could probably beat them in a fight, but antifascists didn't come to outnumber fascists by winning some kind of primordial fistfight between the two sides. They came to outnumber fascists because people rejected fascism on the merits. These merits might not have been "logical" in the sense of Aristotle dispassionately proving lemmas at a chalkboard, but "fascists kill people, killing people is wrong, therefore fascism is wrong" is a sort of folk logical conclusion which is both correct and compelling. Even "a fascist killed my brother, so fuck them" is a placeholder for a powerful philosophical argument making a probabilistic generalization from indexical evidence to global utility. So insofar as violence is asymmetric, it's because it parasitizes on logic which allows the good guys to be more convincing and so field a bigger army. Violence itself doesn't enhance that asymmetry; if anything, it decreases it by giving an advantage to whoever is more ruthless and power-hungry.

The same is true of documentaries. As I said before, Harford can produce as many anti-Trump documentaries as he wants, but Trump can fund documentaries of his own. He has the best documentaries. Nobody has ever seen documentaries like this. They'll be absolutely huge.

And the same is true of rhetoric. Martin Luther King was able to make persuasive emotional appeals for good things. But Hitler was able to make persuasive emotional appeals for bad things. I've previously argued that Mohammed counts as the most successful persuader of all time. These three people pushed three very different ideologies, and rhetoric worked for them all. Robinson writes as if "use rhetoric and emotional appeals" is a novel idea for Democrats, but it seems to me like they were doing little else throughout the election (pieces attacking Trump's character, pieces talking about how inspirational Hillary was, pieces appealing to various American principles like equality, et cetera). It's just that they did a bad job, and Trump did a better one. The real takeaway here is "do rhetoric better than the other guy". But "succeed" is not a primitive action.

Unless you use asymmetric weapons, the best you can hope for is to win by coincidence.

That is, there's no reason to think that good guys are consistently better at rhetoric than bad guys. Some days the Left will have an Obama and win the rhetoric war. Other days the Right will have a Reagan and *they'll* win the rhetoric war. Overall you should average out to a 50% success rate. When you win, it'll be because you got lucky.

And there's no reason to think that good guys are consistently better at documentaries than bad guys. Some days the NIH will spin a compelling narrative and people will smoke less. Other days the tobacco companies will spin a compelling narrative and people will smoke more. Overall smoking will stay the same. And again, if you win, it's because you lucked out into having better videographers or something.

I'm not against winning by coincidence. If I stumbled across Stalin and I happened to have a gun, I would shoot him without worrying about how it's "only by coincidence" that he didn't have the gun instead of me. You should use your symmetric weapons if for no reason other than that the other side's going to use *theirs* and so you'll have a disadvantage if you don't. But you shouldn't confuse it with a long-term solution.

Improving the quality of debate, shifting people's mindsets from transmission to collaborative truth-seeking, is a painful process. It has to be done one person at a time, it only works on people who are already *almost* ready for it, and you will pick up far fewer warm bodies per hour of work than with any of the other methods. But in an otherwise-random world, even a little purposeful action can make a difference. Convincing 2% of people would have flipped three of the last four US presidential elections. And this is a capacity to win-for-reasons-other-than-coincidence that you can't build any other way.

(and my hope is that the people most willing to engage in debate, and the ones most likely to recognize truth when they see it, are disproportionately influential – scientists, writers, and community leaders who have influence beyond their number and can help others see reason in turn)

I worry that I'm not communicating how beautiful and inevitable all of this is. We're surrounded by a a vast confusion, "a darkling plain where ignorant armies clash by night", with one side or another making a temporary advance and then falling back in turn. And in the middle of all of it, there's this gradual capacity-building going on, where what starts off as a hopelessly weak signal gradually builds up strength, until one army starts winning a little more often than chance, then a lot more often, and finally takes the field entirely. Which seems strange, because surely you can't build any complex signal-detection machinery in the middle of all the chaos, surely you'd be shot the moment you left the trenches, but – *your enemies are helping you do it*. Both sides are diverting their artillery from the relevant areas, pooling their resources, helping bring supplies to the engineers, because until the very end they think it's going to ensure *their* final victory and not yours.

You're doing it right under their noses. They might try to ban your documentaries, heckle your speeches, fight your violence Middlebury-student-for-Middlebury-student – but when it comes to the long-term solution to ensure your complete victory, they'll roll down their sleeves, get out their hammers, and build it alongside you.

A parable: Sally is a psychiatrist. Her patient has a strange delusion: that *Sally* is the patient and *he* is the psychiatrist. She would like to commit him and force medication on him, but he is an important politician and if push comes to shove he might be able to commit *her* instead. In desperation, she proposes a bargain: they will *both* take a certain medication. He agrees; from within his delusion, it's the best way for him-the-psychiatrist to cure her-the-patient. The two take their pills at the same time. The medication works, and the patient makes a full recovery.

(well, half the time. The other half, the medication works and *Sally* makes a full recovery.)

V.

Harford's article says that facts and logic don't work on people. The various lefty articles say they merely don't work on Trump supporters, ie 50% of the population.

If you genuinely believe that facts and logic don't work on people, you shouldn't be writing articles with potential solutions. You should be jettisoning everything you believe and entering a state of pure Cartesian doubt, where you try to rederive everything from *cogito ergo sum*.

If you genuinely believe that facts and logic don't work on at least 50% of the population, again, you shouldn't be writing articles with potential solutions. You should be worrying whether you're in that 50%. After all, how did you figure out you aren't? By using facts and logic? *What did we just say?*

Nobody is doing either of these things, so I conclude that they accept that facts can sometimes work. Asymmetric weapons are not a pipe dream. As Gandhi used to say, "If you think the world is all bad, remember that it contains people like you."

You are not completely immune to facts and logic. But you have been wrong about things before. You may be a bit smarter than the people on the other side. You may even be a *lot* smarter. But fundamentally their problems are your problems, and the same kind of logic that convinced you can convince them. It's just going to be a long slog. You didn't develop *your* opinions after a five-minute shouting match. You developed them after years of education and acculturation and engaging with hundreds of books and hundreds of people. Why should they be any different?

You end up believing that the problem is deeper than insufficient documentary production. The problem is that Truth is a weak signal. You're trying to perceive Truth. You would like to hope that the other side is trying to perceive Truth too. But at least one of you is doing it wrong. It seems like perceiving Truth accurately is harder than you thought.

You believe your mind is a truth-sensing instrument that does at least a little bit better than chance. You *have* to believe that, or else what's the point? But it's like one of those physics experiments set up to detect gravitational waves or something, where it has to be in a cavern five hundred feet underground in a lead-shielded chamber atop a gyroscopically stable platform cooled to one degree above absolute zero, trying to detect fluctuations of a millionth of a centimeter. Except you don't have the cavern or the lead or the gyroscope or the coolants. You're on top of an erupting volcano being pelted by meteorites in the middle of a hurricane.

If you study psychology for ten years, you can remove the volcano. If you spend another ten years obsessively checking your performance in various *metis*-intensive domains, you can remove the meteorites. You can never remove the hurricane and you shouldn't try. But if there are a thousand trustworthy people at a thousand different parts of the hurricane, then the stray gusts of wind will cancel out and they can average their readings to get something approaching a signal.

All of this is too slow and uncertain for a world that needs more wisdom *now*. It would be nice to force the matter, to pelt people with speeches and documentaries until they come around. This will work in the short term. In the long term, it will leave you back where you started.

If you want people to be right more often than chance, you have to teach them ways to distinguish truth from falsehood. If this is in the face of enemy action, you will have to teach them so well that they cannot be fooled. You will have to do it person by person until the signal is strong and clear. You will have to raise the sanity waterline. There is no shortcut.

The Ideology Is Not The Movement

I.

Why is there such a strong Sunni/Shia divide?

I know the Comparative Religion 101 answer. The early Muslims were debating who was the rightful caliph. Some of them said Abu Bakr, others said Ali, and the dispute has been going on ever since. On the other hand, that was fourteen hundred years ago, both candidates are long dead, and there's no more caliphate. You'd think maybe they'd let the matter rest.

Sure, the two groups have slightly different hadith and schools of jurisprudence, but how many Muslims even *know* which school of jurisprudence they're supposed to be following? It seems like a pretty minor thing to have centuries of animus over.

And so we return again to Robbers' Cave:

> The experimental subjects — excuse me, "campers" — were 22 boys between 5th and 6th grade, selected from 22 different schools in Oklahoma City, of stable middle-class Protestant families, doing well in school, median IQ 112. They were as well-adjusted and as similar to each other as the researchers could manage.
>
> The experiment, conducted in the bewildered aftermath of World War II, was meant to investigate the causes—and possible remedies—of intergroup conflict. How would they spark an intergroup conflict to investigate? Well, the 22 boys were divided into two groups of 11 campers, and —
>
> — and that turned out to be quite sufficient.
>
> The researchers' original plans called for the experiment to be conducted in three stages. In Stage 1, each group of campers would settle in, unaware of the other group's existence. Toward the end of Stage 1, the groups would gradually be made aware of each other. In Stage 2, a set of contests and prize competitions would set the two groups at odds.
>
> They needn't have bothered with Stage 2. There was hostility almost from the moment each group became aware of the other group's existence: They were using our campground, our baseball diamond. On their first meeting, the two groups began hurling insults. They named themselves the Rattlers and the Eagles (they hadn't needed names when they were the only group on the campground).
>
> When the contests and prizes were announced, in accordance with pre-established experimental procedure, the intergroup rivalry rose to a fever pitch. Good sportsmanship in the contests was evident for the first two days but rapidly disintegrated.
>
> The Eagles stole the Rattlers' flag and burned it. Rattlers raided the Eagles' cabin and stole the blue jeans of the group leader, which they painted orange and carried as a flag the next day, inscribed with the legend "The Last of the Eagles". The Eagles launched a retaliatory raid on the Rattlers, turning over beds, scattering dirt. Then they returned to their cabin where they entrenched and prepared weapons (socks filled with rocks) in case of a return raid. After the

> Eagles won the last contest planned for Stage 2, the Rattlers raided their cabin and stole the prizes. This developed into a fistfight that the staff had to shut down for fear of injury. The Eagles, retelling the tale among themselves, turned the whole affair into a magnificent victory—they'd chased the Rattlers "over halfway back to their cabin" (they hadn't).
>
> Each group developed a negative stereotype of Them and a contrasting positive stereotype of Us. The Rattlers swore heavily. The Eagles, after winning one game, concluded that the Eagles had won because of their prayers and the Rattlers had lost because they used cuss-words all the time. The Eagles decided to stop using cuss-words themselves. They also concluded that since the Rattlers swore all the time, it would be wiser not to talk to them. The Eagles developed an image of themselves as proper-and-moral; the Rattlers developed an image of themselves as rough-and-tough.

If the researchers had decided that the real difference between the two groups was that the Eagles were adherents of Eagleism, which held cussing as absolutely taboo, and the Rattlers adherents of Rattlerism, which held it a holy duty to cuss five times a day – well, that strikes me as the best equivalent to saying that Sunni and Shia differ over the rightful caliph.

II.

Nations, religions, cults, gangs, subcultures, fraternal societies, internet communities, political parties, social movements – these are all really different, but they also have some deep similarities. They're all groups of people. They all combine comradery within the group with a tendency to dislike other groups of the same type. They all tend to have a stated purpose, like electing a candidate or worshipping a deity, but also serve a *very* important role as impromptu social clubs whose members mostly interact with one another instead of outsiders. They all develop an internal culture such that members of the groups often like the same foods, wear the same clothing, play the same sports, and have the same philosophical beliefs as other members of the group – even when there are only tenuous links or no links at all to the stated purpose. They all tend to develop sort of legendary histories, where they celebrate and exaggerate the deeds of the groups' founders and past champions. And they all tend to inspire something like patriotism, where people are proud of their group membership and express that pride through conspicuous use of group symbols, group songs, et cetera. For better or worse, the standard way to refer to this category of thing is "tribe".

Tribalism is potentially present in all groups, but levels differ a lot even in groups of nominally the same type. Modern Belgium seems like an unusually non-tribal nation; Imperial Japan in World War II seems like an unusually tribal one. Neoliberalism and market socialism seem like unusually non-tribal political philosophies; communism and libertarianism seem like unusually tribal ones. Corporations with names like Amalgamated Products Co probably aren't very tribal; charismatic corporations like Apple that become identities for their employees and customers are more so. Cults are maybe the most tribal groups that exist in the modern world, and those Cult Screening Tools make good measures for tribalism as well.

The dangers of tribalism are obvious; for example, fascism is based around dialing a country's tribalism up to eleven, and it ends poorly. If I had written this essay five years ago, it would be be titled "Why Tribalism Is Stupid And Needs To Be Destroyed".

Since then, I've changed my mind. I've found that I enjoy being in tribes as much as anyone else.

Part of this was resolving a major social fallacy I'd had throughout high school and college, which was that the correct way to make friends was to pick the five most interesting people I knew and try to befriend them. This almost never worked and I thought it meant I had terrible social skills. Then I looked at what everyone else was doing, and I found that instead of isolated surgical strikes of friendship, they were forming groups. The band people. The mock trial people. The football team people. The Three Popular Girls Who Went Everywhere Together. Once I tried "falling in with" a group, friendship became much easier and self-sustaining precisely because of all of the tribal development that happens when a group of similar people all know each other and have a shared interest. Since then I've had good luck finding tribes I like and that accept me – the rationalists being the most obvious example, but even interacting with my coworkers on the same hospital unit at work is better than trying to find and cultivate random people.

Some benefits of tribalism are easy to explain. Tribalism intensifies all positive and prosocial feelings within the tribe. It increases trust within the tribe and allows otherwise-impossible forms of cooperation – remember Haidt on the Jewish diamond merchants outcompeting their rivals because their mutual Judaism gave them a series of high-trust connections that saved them costly verification procedures? It gives people a support network they can rely on when their luck is bad and they need help. It lets you "be yourself" without worrying that this will be incomprehensible or offensive to somebody who thinks totally differently from you. It creates an instant densely-connected social network of people who mostly get along with one another. It makes people feel like part of something larger than themselves, which makes them happy and can (provably) improves their physical and mental health.

Others are more complicated. I can just make motions at a feeling that "what I do matters", in the sense that I will probably never be a Beethoven or a Napoleon who is very important to the history of the world as a whole, but I can do things that are important within the context of a certain group of people. All of this is really good for my happiness and mental health. When people talk about how modern society is "atomized" or "lacks community" or "doesn't have meaning", I think they're talking about a lack of tribalism, which leaves people all alone in the face of a society much too big to understand or affect. The evolutionary psychology angle here is too obvious to even be worth stating.

And others are entirely philosophical. I think some people would say that wanting to have a tribe is like wanting to have a family – part of what it means to be human – and demands to justify either are equally wrong-headed.

Eliezer thinks every cause wants to be a cult. I would phrase this more neutrally as "every cause wants to be a tribe". I've seen a lot of activities go through the following cycle:

1. Let's get together to do X
2. Let's get together to do X, and have drinks afterwards
3. Let's get together to discuss things from an X-informed perspective
4. Let's get together to discuss the sorts of things that interest people who do X
5. Let's get together to discuss how the sort of people who do X are much better than the sort of people who do Y.
6. Dating site for the sort of people who do X

7. Oh god, it was so annoying, she spent the whole date talking about X.
8. X? What X?

This can happen over anything or nothing at all. Despite the artificial nature of the Robbers' Cove experiment, its groups are easily recognized as tribes. Indeed, the reason this experiment is so interesting is that it shows tribes in their purest form; no veneer of really being about pushing a social change or supporting a caliph, just tribes for tribalism's sake.

III.

Scholars call the process of creating a new tribe "ethnogenesis" – Robbers' Cave was artificially inducing ethnogenesis to see what would happen. My model of ethnogenesis involves four stages: pre-existing differences, a rallying flag, development, and dissolution.

Pre-existing differences are the raw materials out of which tribes are made. A good tribe combines people who have similar interests and styles of interaction *even before* the ethnogenesis event. Any description of these differences will necessarily involve stereotypes, but a lot of them should be hard to argue. For example, atheists are often pretty similar to one another even before they deconvert from their religion and officially become atheists. They're usually nerdy, skeptical, rational, not very big on community or togetherness, sarcastic, well-educated. At the risk of going into touchier territory, they're pretty often white and male. You take a sample of a hundred equally religious churchgoers and pick out the ones who are *most like the sort of people who are atheists* even if all of them are 100% believers. But there's also something more than that. There are subtle habits of thought, not yet described by any word or sentence, which atheists are more likely to have than other people. It's part of the reason why atheists *need* atheism as a rallying flag instead of just starting the Skeptical Nerdy Male Club.

The rallying flag is the explicit purpose of the tribe. It's usually a belief, event, or activity that get people with that specific pre-existing difference together and excited. Often it brings previously latent differences into sharp relief. People meet around the rallying flag, encounter each other, and say "You seem like a kindred soul!" or "I thought I was the only one!" Usually it suggests some course of action, which provides the tribe with a purpose. For atheists, the rallying flag is not believing in God. Somebody says "Hey, I don't believe in God, if you also don't believe in God come over here and we'll hang out together and talk about how much religious people suck." All the atheists go over by the rallying flag and get very excited about meeting each other. It starts with "Wow, you hate church too?", moves on to "Really, you also like science fiction?", and ends up at "Wow, you have the same undefinable habits of thought that I do!"

Development is all of the processes by which the fledgling tribe gains its own culture and history. It's a turning-inward and strengthening-of-walls, which transforms it from 'A Group Of People Who Do Not Believe In God And Happen To Be In The Same Place' to 'The Atheist Tribe'. For example, atheists have symbols like that 'A' inside an atom. They have jokes and mascots like Russell's Teapot and the Invisible Pink Unicorn. They have their own set of heroes, both mythologized past heroes like Galileo and controversial-but-undeniably-important modern heroes like Richard Dawkins and Daniel Dennett. They have celebrities like P.Z. Myers and Hemant Mehta. They have universally-agreed-upon villains to be booed and hated, like televangelists or the Westboro Baptist Church. They have grievances, like all the times that atheists have

been fired or picked on by religious people, and all the laws about pledging allegiance to one nation under God and so on. They have stereotypes about themselves – intelligent, helpful, passionate – and stereotypes about their outgroups – deluded, ignorant, bigoted.

Dissolution is optional. The point of the previous three steps is to build a "wall" between the tribe and the outside, a series of systematic differences that let everybody know which side they're on. If a tribe was never really that different from the surrounding population, stops caring that much about its rallying flag, and doesn't develop enough culture, then the wall fails and the members disperse into the surrounding population. The classic example is the assimilation of immigrant groups like Irish-Americans, but history is littered with failed communes, cults, and political movements. Atheism hasn't quite dissolved yet, but occasionally you see hints of the process. A lot of the comments around "Atheism Plus" centered around this idea of "Okay, talking about how there's no God all the time has gotten boring, plus nobody interesting believes in God anymore anyway, so let's become about social justice instead". The parts of atheism who went along with that message mostly dissolved into the broader social justice community – there are a host of nominally atheist blogs that haven't talked about anything except social justice in months. Other fragments of the atheist community dissolved into transhumanism, or libertarianism, or any of a number of other things. Although there's still an atheist community, it no longer seems quite as vibrant and cohesive as it used to be.

We can check this four-stage model by applying it to the Sunni and Shia and seeing if it sticks.

I know very little about early Islam and am relying on sources that might be biased, so don't declare a fatwa against me if I turn out to be wrong, but it looks like from the beginning there were big pre-existing differences between proto-Shia and proto-Sunni. A lot of Ali's earliest supporters were original Muslims who had known Mohammed personally, and a lot of Abu Bakr's earliest supporters were later Muslims high up in the Meccan/Medinan political establishment who'd converted only after it became convenient to do so. It's really easy to imagine cultural, social, and personality differences between these two groups. Probably members in each group already knew one another pretty well, and already had ill feelings towards members of the other, without necessarily being able to draw the group borders clearly or put their exact differences into words. Maybe it was "those goody-goodies who are always going on about how close to Mohammed they were but have no practical governing ability" versus "those sellouts who don't really believe in Islam and just want to keep playing their political games".

Then came the rallying flag: a political disagreement over the succession. One group called themselves "the party of Ali", whose Arabic translation "Shiatu Ali" eventually ended up as just "Shia". The other group won and called itself "the traditional orthodox group", in Arabic "Sunni". Instead of a vague sense of "I wonder whether that guy there is one of those goody-goodies always talking about Mohammed, or whether he's a practical type interested in good governance", people could just ask "Are you for Abu Bakr or Ali?" and later "Are you Sunni or Shia?" Also at some point, I'm not exactly sure how, most of the Sunni ended up in Arabia and most of the Shia ended up in Iraq and Iran, after which I think some pre-existing Iraqi/Iranian vs. Arab cultural differences got absorbed into the Sunni/Shia mix too.

Then came development. Both groups developed elaborate mythologies lionizing their founders. The Sunni got the history of the "rightly-guided caliphs", the Shia

exaggerated the first few imams to legendary proportions. They developed grievances against each other; according to Shia history, the Sunnis killed eleven of their twelve leaders, with the twelfth escaping only when God directly plucked him out of the world to serve as a future Messiah. They developed different schools of hadith interpretation and jurisprudence and debated the differences ad nauseum with each other for hundreds of years. A lot of Shia theology is in Farsi; Sunni theology is entirely in Arabic. Sunni clergy usually dress in white; Shia clergy usually dress in black and green. Not all of these were deliberately done in opposition to one another; most were just a consequence of the two camps being walled off from one another and so allowed to develop cultures independently.

Obviously the split hasn't dissolved yet, but it's worth looking at similar splits that have. Catholicism vs. Protestantism is still a going concern in a few places like Ireland, but it's nowhere near the total wars of the 17th century or even the Know-Nothing-Parties of the 19th. Consider that Marco Rubio is Catholic, but nobody except Salon particularly worries about that or says that it will make him unsuitable to lead a party representing the interests of very evangelical Protestants. Heck, the same party was happy to nominate Mitt Romney, a Mormon, and praise him for his "Christian faith". Part of it is the subsumption of those differences into a larger conflict – most Christians acknowledge Christianity vs. atheism to be a bigger deal than interdenominational disputes these days – and part of it is that everyone of every religion is so influenced by secular American culture that the religions have been reduced to their rallying flags alone rather than being fully developed tribes at this point. American Sunni and Shia seem to be well on their way to dissolving into each other too.

IV.

I want to discuss a couple of issues that I think make more sense once you understand the concept of tribes and rallying flags:

1. Disability: I used to be very confused by disabled people who insist on not wanting a "cure" for their condition. Deaf people and autistic people are the two classic examples, and sure enough we find articles like Not All Deaf People Want To Be Cured and They Don't Want An Autism Cure. Autistic people can at least argue their minds work differently rather than worse, but being deaf seems to be a straight-out disadvantage: the hearing can do anything the deaf can, and can hear also. A hearing person can become deaf at any time just by wearing earplugs, but a deaf person can't become hearing, at least not without very complicated high-tech surgeries.

When I asked some deaf friends about this, they explained that they had a really close-knit and supportive deaf culture, and that most of their friends, social events, and ways of relating to other people and the world were through this culture. This made sense, but I always wondered: if you were able to hear, couldn't you form some other culture? If worst came to worst and nobody else wanted to talk to you, couldn't you at least have the Ex-Deaf People's Club?

I don't think so. Deafness acts as a rallying flag that connects people, gives them a shared foundation to build culture off of, and walls the group off from other people. If all deaf people magically became able to hear, their culture would eventually drift apart, and they'd be stuck without an ingroup to call their own.

Part of this is reasonable cost-benefit calculation – our society is so vast and atomized, and forming real cohesive tribes is so hard, that they might reasonably expect it would

be a lot of trouble to find another group they liked as much as the deaf community. But another part of this seems to be about an urge to cultural self-preservation.

2. Genocide: This term is kind of overused these days. I always thought of it as meaning literally killing every member of a certain group – the Holocaust, for example – but the new usage includes "cultural genocide". For example, autism rights advocates sometimes say that anybody who cured autism would be committing genocide – this is of course soundly mocked, but it makes sense if you think of autistic people as a tribe that would be dissolved absent its rallying flag. The tribe would be eliminated – thus "cultural genocide" is a reasonable albeit polemical description.

It seems to me that people have an urge toward cultural self-preservation which is as strong or stronger as the urge to individual self-preservation. Part of this is rational cost-benefit calculation – if someone loses their only tribe and ends up alone in the vast and atomized sea of modern society, it might take years before they can find another tribe and really be at home there. But a lot of it seems to be beyond that, an emotional certainty that losing one's culture and having it replaced with another is not okay, any more than being killed at the same time someone else has a baby is okay. Nor do I think this is necessarily irrational; locating the thing whose survival you care about in the self rather than the community is an assumption, and people can make different assumptions without being obviously wrong.

3. Rationalists: The rationalist community is a group of people (of which I'm a part) who met reading the site Less Wrong and who tend to hang out together online, sometimes hang out together in real life, and tend to befriend each other, work with each other, date each other, and generally move in the same social circles. Some people call it a cult, but that's more a sign of some people having lost vocabulary for anything between "totally atomized individuals" and "outright cult" than any particular cultishness.

But people keep asking me what exactly the rationalist community *is*. Like, what is the thing they believe that makes them rationalists? It can't just be about being rational, because loads of people are interested in that and most of them aren't part of the community. And it can't just be about transhumanism because there are a lot of transhumanists who aren't rationalists, and lots of rationalists who aren't transhumanists. And it can't just be about Bayesianism, because pretty much everyone, rationalist or otherwise, agrees that is a kind of statistics that is useful for some things but not others. So what, exactly, is it?

This question has always bothered me, but now after thinking about it a lot I finally have a clear answer: rationalism is the belief that Eliezer Yudkowsky is the rightful caliph.

No! Sorry! I think "the rationalist community" is a tribe much like the Sunni or Shia that started off with some pre-existing differences, found a rallying flag, and then developed a culture.

The pre-existing differences range from the obvious to the subtle. A lot of rationalists are mathematicians, programmers, or computer scientists. The average IQ is in the 130s. White men are overrepresented, but so are LGBT and especially transgender people. But there's more. Nobody likes the Myers-Briggs test, but I continue to find it really interesting that rationalists have some Myers-Briggs types (INTJ/INTP) at ten times the ordinary rate, and other types (ISFJ/ESFP) at only one one-hundredth the ordinary rate. Myers-Briggs doesn't cleave reality at its joints, but if it measures

anything at all about otherwise hard-to-explain differences in thinking styles, the rationalist community heavily selects for those same differences. Sure enough, I am *constantly* running into people who say "This is the only place where I've ever found people who think like me" or "I finally feel understood".

The rallying flag was the Less Wrong Sequences. Eliezer Yudkowsky started a blog (actually, borrowed Robin Hanson's) about cognitive biases and how to think through them. Whether or not you agreed with him or found him enlightening loaded heavily on those pre-existing differences, so the people who showed up in the comment section got along and started meeting up with each other. "Do you like Eliezer Yudkowsky's blog?" became a useful proxy for all sorts of things, eventually somebody coined the word "rationalist" to refer to people who did, and then you had a group with nice clear boundaries.

The development is everything else. Obviously a lot of jargon sprung up in the form of terms from the blog itself. The community got heroes like Gwern and Anna Salamon who were notable for being able to approach difficult questions insightfully. It doesn't have much of an outgroup yet – maybe just bioethicists and evil robots. It has its own foods – MealSquares, that one kind of chocolate everyone in Berkeley started eating around the same time – and its own games. It *definitely* has its own inside jokes. I think its most important aspect, though, is a set of shared mores – everything from "understand the difference between ask and guess culture and don't get caught up in it" to "cuddling is okay" to "don't misgender trans people" – and a set of shared philosophical assumptions like utilitarianism and reductionism.

I'm stressing this because I keep hearing people ask "What is the rationalist community?" or "It's really weird that I seem to be involved in the rationalist community even though I don't share belief X" as if there's some sort of necessary-and-sufficient featherless-biped-style ideological criterion for membership. This is why people are saying "Lots of you aren't even singularitarians, and everyone agrees Bayesian methods are useful in some places and not so useful in others, so what is your community even *about*?" But once again, it's ~~about Eliezer Yudkowsky being the rightful caliph~~ it's not necessarily *about* anything.

If you take only one thing from this essay, it's that communities are best understood not logically but historically. If you want to understand the Shia, don't reflect upon the true meaning of Ali being the rightful caliph, understand that a dispute involving Ali initiated ethnogenesis, the resulting culture picked up a bunch of features and became useful to various people, and now here we are. If you want to understand the rationalist community, don't ask exactly how near you have to think the singularity has to be before you qualify for membership, focus on the fact that some stuff Eliezer Yudkowsky wrote led to certain people identifying themselves as "rationalists" and for various reasons I enjoy dinner parties with those people about 10000% more interesting than dinner parties with randomly selected individuals.

nostalgebraist actually summed this up really well: "Maybe the real rationalism was the friends we made along the way." Maybe that's the real Shia Islam too, and the real Democratic Party, and so on.

4. Evangelical And Progressive Religion: There seems to be a generational process, sort of like Harold Lee's theory of immigrant assimilation, by which religions dissolve. The first generation believes everything literally. The second generation believes that the religion might not be literally true, but it's an important expression of universal values and they still want to follow the old ways and participate in the

church/temple/mosque/mandir community. The third generation is completely secularized.

This was certainly my family's relationship with Judaism. My great-great-grandfather was so Jewish that he left America and returned to Eastern Europe because he was upset at American Jews for not being religious enough. My great-grandfather stayed behind in America but remained a very religious Jew. My grandparents attend synagogue when they can remember, speak a little Yiddish, and identify with the traditions. My parents went to a *really* liberal synagogue where the rabbi didn't believe in God and everyone just agreed they were going through the motions. I got Bar Mitzvahed when I was a kid but haven't been to synagogue in years. My children probably won't even have that much.

So imagine you're an evangelical Christian. All the people you like are also evangelical Christians. Most of your social life happens at church. Most of your good memories involve things like Sunday school and Easter celebrations, and even your bittersweet memories are things like your pastor speaking at your parents' funeral. Most of your hopes and dreams involve marrying someone and having kids and then sharing similarly good times with them. When you try to hang out with people who aren't evangelical Christians, they seem to think really differently than you do, and not at all in a good way. A lot of your happiest intellectual experiences involve geeking out over different Bible verses and the minutiae of different Christian denominations.

Then somebody points out to you that God probably doesn't exist. And even if He does, it's probably in some vague and complicated way, and not the way that means that the Thrice-Reformed Meta-Baptist Church and *only* the Thrice-Reformed Meta-Baptist Church has the correct interpretation of the Bible and everyone else is wrong.

On the one hand, their argument might be convincing. On the other, you are pretty sure that if everyone agreed on this, your culture would be destroyed. Sure, your kids could be Christmas-and-Easter-Christians who still enjoy the cultural aspects and derive personal meaning from the Bible. But you're pretty sure that within a couple of generations your descendents would be exactly as secular as anyone else. Absent the belief that serves as your culture's wall against the outside world, it would dissolve without a trace into the greater homogeneity of Western liberal society. So, do you keep believing a false thing? Or do you give up on everything you love and enjoy and dissolve into a culture that mostly hates and mocks people like you? There's no good choice. This is why it sucks that things like religion and politics are both rallying flags for tribes, and actual things that there may be a correct position on.

5. Religious Literalism: One comment complaint I heard during the height of the Atheist-Theist Online Wars was that atheists were a lot like fundamentalists. Both wanted to interpret the religious texts in the most literal possible way.

Being on the atheist side of these wars, I always wanted to know: well, why wouldn't you? Given that the New Testament clearly says you have to give all your money to the poor, and the Old Testament doesn't say anything about mixing meat and milk, maybe religious Christians should start giving everything to the poor and religious Jews should stop worrying so much about which dishes to use when?

But I think this is the same mistake as treating the Sunni as an organization dedicated to promoting an Abu Bakr caliphate. The holy book is the rallying flag for a religion, but the religion is not itself about the holy book. The rallying flag created a walled-off space where people could undergo the development process and create an

independent culture. That independent culture may diverge significantly from the holy book.

I think that very neurotypical people naturally think in terms of tribes, and the idea that they have to retool their perfectly functional tribe to conform to the exact written text of its holy book or constitution or stated political ideology or something seems silly to them. I think that less neurotypical people – a group including many atheists – think less naturally in terms of tribes and so tend to take claims like "Christianity is about following the Bible" at face value. But Christianity is about being part of the Christian tribe, and although that tribe started around the Bible, maintains its coherence because of the Bible, and is of course naturally influenced by it, if it happens to contradict the Bible in some cases that's not necessarily surprising or catastrophic.

This is also why I'm not really a fan of debates over whether Islam is really "a religion of peace" or "a religion of violence", especially if those debates involve mining the Quran for passages that support one's preferred viewpoint. It's not just because the Quran is a mess of contradictions with enough interpretive degrees of freedom to prove anything at all. It's not even because Islam is a host of separate cultures as different from one another as Unitarianism is from the Knights Templar. It's because the Quran just created the space in which the Islamic culture could evolve, but had only limited impact on that evolution. As well try to predict the warlike or peaceful nature of the United Kingdom by looking at a topographical map of Great Britain.

6. Cultural Appropriation: Thanks to some people who finally explained this to me in a way that made sense. When an item or artform becomes the rallying flag for a tribe, it can threaten the tribe if other people just want to use it as a normal item or artform.

Suppose that rappers start with pre-existing differences from everyone else. Poor, male, non-white minority, lots of experience living in violent places, maybe a certain philosophical outlook towards their condition. Then they get a rallying flag: rap music. They meet one another, like one another. The culture undergoes further development: the lionization of famous rappers, the development of a vocabulary of shared references. They get all of the benefits of being in a tribe like increased trust, social networking, and a sense of pride and identity.

Now suppose some rich white people get into rap. Maybe they get into rap for innocuous reasons: rap is cool, they like the sound of it. Fine. But they don't share the pre-existing differences, and they can't be easily assimilated into the tribe. Maybe they develop different conventions, and start saying that instead of being about the struggles of living in severe poverty, rap should be about Founding Fathers. Maybe they start saying the original rappers are bad, and they should stop talking about violence and bitches because that ruins rap's reputation. Since rich white people tend to be be good at gaining power and influence, maybe their opinions are overrepresented at the Annual Rap Awards, and all of a sudden you can't win a rap award unless your rap is about the Founding Fathers and doesn't mention violence (except Founding-Father-related duels). All of a sudden if you try to start some kind of impromptu street rap-off, you're no longer going to find a lot of people like you whom you instantly get along with and can form a high-trust community. You're going to find half people like that, and half rich white people who strike you as annoying and are always complaining that your raps don't feature any Founding Fathers at all. The rallying flag fails and the tribe is lost as a cohesive entity.

7. Fake Gamer Girls: A more controversial example of the same. Video gaming isn't just a fun way to pass the time. It also brings together a group of people with some pre-existing common characteristics: male, nerdy, often abrasive, not very successful, interested in speculation, high-systematizing. It gives them a rallying flag and creates a culture which then develops its own norms, shared reference points, internet memes, webcomics, heroes, shared gripes, even some unique literature. Then other people with very different characteristics and no particular knowledge of the culture start enjoying video games just because video games are fun. Since the Gamer Tribe has no designated cultural spaces except video games forums and magazines, they view this as an incursion into their cultural spaces and a threat to their existence as a tribe.

Stereotypically this is expressed as them getting angry when girls start playing video games. One can argue that it's unfair to infer tribe membership based on superficial characteristics like gender – in the same way it might be unfair for the Native Americans to assume someone with blonde hair and blue eyes probably doesn't follow the Old Ways – but from the tribe's perspective it's a reasonable first guess.

I've found gamers to get along pretty well with women who share their culture, and poorly with men who don't – but admit that the one often starts from an assumption of foreignness and the other from an assumption of membership. More important, I've found the *idea* of the rejection of the 'fake gamer girl', real or not, raised more as a libel by people who genuinely *do* want to destroy gamer culture, in the sense of cleansing video-game-related spaces of a certain type of person/culture and making them entirely controlled by a different type of person/culture, in much the same way that a rich white person who says any rapper who uses violent lyrics needs to be blacklisted from the rap world has a clear culture-change project going on.

These cultural change projects tend to be framed in terms of which culture has the better values, which I think is a limited perspective. I think America has better values than Pakistan does, but that doesn't mean I want us invading them, let alone razing their culture to the ground and replacing it with our own.

8. Subcultures And Posers: Obligatory David Chapman link. A poser is somebody who uses the rallying flag but doesn't have the pre-existing differences that create tribal membership and so never really fits into the tribe.

9. Nationalism, Patriotism, and Racism: Nationalism and patriotism use national identity as the rallying flag for a strong tribe. In many cases, nationalism becomes ethno-nationalism, which builds tribal identity off of a combination of heritage, language, religion, and culture. It has to be admitted that this can make for some *incredibly* strong tribes. The rallying flag is built into ancestry, and so the walls are near impossible to obliterate. The symbolism and jargon and cultural identity can be instilled from birth onward. Probably the best example of this is the Jews, who combine ethnicity, religion, and language into a bundle deal and have resisted assimilation for millennia.

Sometimes this can devolve into racism. I'm not sure exactly what the difference between ethno-nationalism and racism is, or whether there even *is* a difference, except that "race" is a much more complicated concept than ethnicity and it's probably not a coincidence that it has become most popular in a country like America whose ethnicities are hopelessly confused. The Nazis certainly needed a lot of work to transform concern about the German nation into concern about the Aryan race. But it's fair to say all of this is somewhat related or at least potentially related.

On the other hand, in countries that have non-ethnic notions of heritage, patriotism has an opportunity to substite for racism. Think about the power of the civil rights message that, whether black or white, we are all Americans.

This is maybe most obvious in sub-national groups. Despite people paying a lot of attention to the supposed racism of Republicans, the rare black Republicans do shockingly well within their party. Both Ben Carson and Herman Cain briefly topped the Republican presidential primary polls during their respective election seasons, and their failures seem to have had much more to do with their own personal qualities than with some sort of generic Republican racism. I see the same with Thomas Sowell, with Hispanic Republicans like Ted Cruz, and Asian Republicans like Bobby Jindal.

Maybe an even stronger example is the human biodiversity movement, which many people understandably accuse of being entirely about racism. Nevertheless, some of its most leading figures are black – JayMan and Chanda Chisala (who is adjacent to the movement but gets lots of respect within it) – and they seem to get equal treatment and respect to their white counterparts. Their membership in a strong and close-knit tribe screens off everything else about them.

I worry that attempts to undermine nationalism/patriotism in order to fight racism risk backfiring. The weaker the "American" tribe becomes, the more people emphasize their other tribes – which can be either overtly racial or else heavily divided along racial lines (eg political parties). It continues to worry me that people who would never display an American flag on their lawn because "nations are just a club for hating foreigners" now have a campaign sign on their lawn, five bumper stickers on their car, and are identifying more and more strongly with political positions – ie clubs for hating their fellow citizens.

Is there such a thing as conservation of tribalism? Get rid of one tribal identity and people just end up seizing on another? I'm not sure. And anyway, nobody can agree on exactly what the American identity or American tribe is anyway, so any conceivable such identity would probably risk alienating a bunch of people. I guess that makes it a moot point. But I still think that deliberately trying to eradicate patriotism is not as good an idea as is generally believed.

V.

I think tribes are interesting and underdiscussed. And in a lot of cases when they are discussed, it's within preexisting frameworks that tilt the playing field towards recognizing some tribes as fundamentally good, others as fundamentally bad, and ignoring the commonalities between all of them.

But in order to talk about tribes coherently, we need to talk about rallying flags. And that involves admitting that a lot of rallying flags are based on ideologies (which are sometimes wrong), holy books (which are *always* wrong), nationality (which we can't define), race (which is racist), and works of art (which some people inconveniently want to enjoy just as normal art without any connotations).

My title for this post is also my preferred summary: the ideology is not the movement. Or, more jargonishly – the rallying flag is not the tribe. People are just trying to find a tribe for themselves and keep it intact. This often involves defending an ideology they might not be tempted to defend for any other reason. This doesn't make them bad, and it *may* not even necessarily mean their tribe deserves to go extinct. I'm reluctant to say for sure whether I think it's okay to maintain a tribe based on a faulty ideology,

but I think it’s at least important to understand that these people are in a crappy situation with no good choices, and they deserve some pity.

Some vital aspects of modern society – freedom of speech, freedom of criticism, access to multiple viewpoints, the existence of entryist tribes with explicit goals of invading and destroying competing tribes as problematic, and the overwhelming pressure to dissolve into the Generic Identity Of Modern Secular Consumerism – make maintaining tribal identities really hard these days. I think some of the most interesting sociological questions revolve around whether there are any ways around the practical and moral difficulties with tribalism, what social phenomena are explicable as the struggle of tribes to maintain themselves in the face of pressure, and whether tribalism continues to be a worthwhile or even a possible project at all.

EDIT: I’ve been informed of a very similar Melting Asphalt post, Religion Is Not About Beliefs. Everyone has pre-stolen my best ideas 😕

Archipelago and Atomic Communitarianism

I.

In the old days, you had your Culture, and that was that. Your Culture told you lots of stuff about what you were and weren't allowed to do, and by golly you listened. Your Culture told you to work the job prescribed to you by your caste and gender, to marry who your parents told you to marry or at *least* someone of the opposite sex, to worship at the proper temples and the proper times, and to talk about *proper* things as opposed to the blasphemous things said by the tribe over there.

Then we got Liberalism, which said all of that was mostly bunk. Like Wicca, its motto is "Do as you will, so long as it harms none". Or in more political terms, "Your right to swing your fist ends where my nose begins" or "If you don't like gay sex, don't have any" or "If you don't like this TV program, don't watch it" or "What happens in the bedroom between consenting adults is none of your business" or "It neither breaks my arm nor picks my pocket". Your job isn't to enforce your conception of virtue upon everyone to build the Virtuous Society, it's to live your own life the way you want to live it and let other people live *their* own lives the way *they* want to live them. This is the much-maligned "atomic individualism," or maybe just liberalism boiled down to its pure essence.

But atomic individualism wasn't as great a solution as it sounded. Maybe one of the first cracks was tobacco ads. Even though putting up a billboard saying "SMOKE MARLBORO" neither breaks anyone's arm nor picks their pocket, it shifts social expectations in such a way that bad effects occur. It's hard to dismiss that with "Well, it's people's own choice to smoke and they should live their lives the way they want" if studies show that more people will want to live their lives in a way that gives them cancer in the presence of the billboard than otherwise.

From there we go into policies like Michael Bloomberg's ban on giant sodas. While the soda ban itself was probably as much symbolic as anything, it's hard to argue with the impetus behind it – a culture where everyone gets exposed to the option to buy very very unhealthy food all the time is going to be less healthy than one where there are some regulations in place to make EAT THIS DONUT NOW a less salient option. I mean, I *know* this is true. A few months ago when I was on a diet I *cringed* every time one my coworkers brought in a box of free donuts and placed wide-open in the doctors' lounge; there was *no way* I wasn't going to take one (or two, or three). I could ask people to stop, but they probably wouldn't, and even if they did I'd just encounter the wide-open box of free donuts *somewhere else*. I'm not proposing that it is *ethically wrong* to bring in free donuts or that banning them is the correct policy, but I do want to make it clear that stating "it's your free choice to partake or not" doesn't eliminate the problem, and that this points to an entire class of serious issues where atomic individualism as construed above is at best an imperfect heuristic.

And I would be remiss talking about the modern turn away from individualism without mentioning social justice. The same people who once deployed individualistic arguments against conservatives: "If you don't like profanity, don't use it", "If you don't like this offensive TV show, don't watch it", "If you don't like pornography, don't buy it" – are now concerned about people using ethnic slurs, TV shows without enough

minority characters, and pornography that encourages the objectification of women. I've objected to some of this on purely empirical grounds, but the least convenient possible world is the one where the purely empirical objections fall flat. If they ever discover proof positive that yeah, pornographication makes women hella objectified, is it acceptable to censor or ban misogynist media on a society-wide level?

And if the answer is yes – and if such media like really, *really* increases the incidence of rape I'm not sure how it couldn't be – then what about all those conservative ideas we've been neglecting for so long? What if strong, cohesive, religious, demographically uniform communities make people more trusting, generous, and cooperative in a way that *also* decreases violent crime and other forms of misery? We have lots of evidence that this is true, and although we can doubt each individual study, we owe conservatives the courtesy of imagining the possible world in which they are right, the same as anti-misogyny leftists. Maybe media glorifying criminals or lionizing nonconformists above those who quietly follow cultural norms has the same kind of erosive effects on "values" as misogynist media. Or, at the very least, we ought to have a good philosophy in place so that we have some idea what to do it if does.

II.

A while ago, in Part V of this essay, I praised liberalism as the only peaceful answer to Hobbes' dilemma of the war of all against all.

Hobbes says that if everyone's fighting then everyone loses out. Even the winners probably end up worse off than if they had just been able to live in peace. He says that governments are good ways to prevent this kind of conflict. Someone – in his formulation a king – tells everyone else what they're going to do, and then everyone else does it. No fighting necessary. If someone tries to start a conflict by ignoring the king, the king crushes them like a bug, no prolonged fighting involved.

But this replaces the problem of potential warfare with the problem of potential tyranny. So we've mostly shifted from absolute monarchies to other forms of government, which is all nice and well except that governments allow a *different* kind of war of all against all. Instead of trying to kill their enemies and steal their stuff, people are tempted to ban their enemies and confiscate their stuff. Instead of killing the Protestants, the Catholics simply ban Protestantism. Instead of forming vigilante mobs to stone homosexuals, the straights merely declare homosexuality is punishable by death. It *might* be better than the alternative – at least everyone knows where they stand and things stay peaceful – but the end result is still a lot of pretty miserable people.

Liberalism is a new form of Hobbesian equilibrium where the government enforces not only a ban on killing and stealing from people you don't like, but also a ban on tyrannizing them out of existence. This is the famous "freedom of religion" and "freedom of speech" and so on, as well as the "freedom of what happens in the bedroom between consenting adults". The Catholics don't try to ban Protestantism, the Protestants don't try to ban Catholicism, and everyone is happy.

Liberalism only works when it's clear to everyone on all sides that there's a certain neutral principle everyone has to stick to. The neutral principle can't be the Bible, or Atlas Shrugged, or anything that makes it look like one philosophy is allowed to judge the others. Right now that principle is the Principle of Harm: you can do whatever you like unless it harms other people, in which case stop. We seem to have inelegantly

tacked on an “also, we can collect taxes and use them for a social safety net and occasional attempts at social progress”, but it seems to be working pretty okay too.

The Strict Principle of Harm says that pretty much the only two things the government can get angry at is literally breaking your leg or picking your pocket – violence or theft. The Loose Principle of Harm says that the government can get angry at complicated indirect harms, things that Weaken The Moral Fabric Of Society. Like putting up tobacco ads. Or having really really big sodas. Or publishing hate speech against minorities. Or eroding trust in the community. Or media that objectifies women.

No one except the most ideologically pure libertarians seems to want to insist on the Strict Principle of Harm. But allowing the Loose Principle Of Harm restores all of the old wars to control other people that liberalism was supposed to prevent. The one person says “Gay marriage will result in homosexuality becoming more accepted, leading to increased rates of STDs! That’s a harm! We must ban gay marriage!” Another says “Allowing people to send their children to non-public schools could lead to kids at religious schools that preach against gay people, causing those children to commit hate crimes when they grow up! That’s a harm! We must ban non-public schools!” And so on, forever.

And I’m talking about non-governmental censorship just as much as government censorship. Even in the most anti-gay communities in the United States, the laws usually allow homosexuality or oppose it only in very weak, easily circumvented ways. The real problem for gays in these communities is the social pressure – whether that means disapproval or risk of violence – that they would likely face for coming out. This too is a violation of liberalism, and it’s one that’s as important or more important than the legal sort.

And right now our way of dealing with these problems is to argue them. “Well, gay people don’t really increase STDs too much.” Or “Home-schooled kids do better than public-schooled kids, so we need to allow them.” The problem is that arguments never terminate. Maybe if you’re *incredibly* lucky, after years of fighting you can get a couple of people on the other side to admit your side is right, but this is a pretty hard process to trust. The great thing about religious freedom is that it short-circuits the debate of “Which religion is correct, Catholicism or Protestantism?” and allows people to tolerate both Catholics and Protestants even if they are divided about the answer to this object-level question. The great thing about freedom of speech is that it short-circuits the debate of “Which party is correct, the Democrats or Republicans?” and allows people to express both liberal and conservative opinions even if they are divided about the object-level question.

If we force all of our discussions about whether to ban gay marriage or allow home schooling to depend on resolving the dispute about whether they indirectly harm the Fabric of Society in some way, we’re forcing dependence on object-level arguments in a way that historically has been very very bad.

Presumably here the more powerful groups would win out and be able to oppress the less powerful groups. We end up with exactly what liberalism tried to avoid – a society where everyone is the guardian of the virtue of everyone else, and anyone who wants to live their lives in a way different from the community’s consensus is out of luck.

In Part I, I argued that *not allowing* people to worry about culture and community at all was inadequate, because these things really do matter.

Here I'm saying that if we *do allow* people to worry about culture and community, we risk the bad old medieval days where all nonconformity gets ruthlessly quashed.

Right now we're balanced precariously between the two states. There's a lot of liberalism, and people are generally still allowed to be gay or home-school their children or practice their religion or whatever. But there's also quite a bit of Enforced Virtue, where kids are forbidden to watch porn and certain kinds of media are censored and in some communities mentioning that you're an atheist will get you Dirty Looks.

It tends to work okay for most of the population. Better than the alternatives, maybe? But there's still a lot of the population that's not free to do things that are very important to them. And there's also a lot of the population that would like to live in more "virtuous" communities, whether it's to lose weight faster or avoid STDs or not have to worry about being objectified. Dealing with these two competing issues is a pretty big part of political philosophy and one that most people don't have any principled solution for.

III.

Imagine a new frontier suddenly opening. Maybe a wizard appears and gives us a map to a new archipelago that geographers had missed for the past few centuries. He doesn't want to rule the archipelago himself, though he will reluctantly help kickstart the government. He just wants to give directions and a free galleon to anybody who wants one and can muster a group of likeminded friends large enough to start a self-sustaining colony.

And so the equivalent of our paleoconservatives go out and found communities based on virtue, where all sexual deviancy is banned and only wholesome films can be shown and people who burn the flag are thrown out to be eaten by wolves.

And the equivalent of our social justiciars go out and found communities where all movies have to have lots of strong minority characters in them, and all slurs are way beyond the pale, and nobody misgenders anybody.

And the equivalent of our Objectivists go out and found communities based totally on the Strict Principle of Harm where everyone is allowed to do whatever they want and there are no regulations on business and everything is super-capitalist all the time.

And some people who just really want to lose weight go out and found communities where you're not allowed to place open boxes of donuts in the doctors' lounge.

Usually the communities are based on a charter, which expresses some founding ideals and asks only the people who agree with those ideals to enter. The charter also specifies a system of government. It could be an absolute monarch, charged with enforcing those ideals upon a population too stupid to know what's good for them. Or it could be a direct democracy of people who all agree on some basic principles but want to work out for themselves what direction the principles take them.

After a while the wizard decides to formalize and strengthen his system, not to mention work out some of the ethical dilemmas.

First he bans communities from declaring war on each other. That's an *obvious* gain. He could just smite warmongers, but he thinks it's more natural and organic to get all the communities into a united government (UniGov for short). Every community

donates a certain amount to a military, and the military's only job is to quash anyone from any community who tries to invade another.

Next he addresses externalities. For example, if some communities emit a lot of carbon, and that causes global warming which threatens to destroy other communities, UniGov puts a stop to that. If the offending communities refuse to stop emitting carbon, then there's that military again.

The third thing he does is prevent memetic contamination. If one community wants to avoid all media that objectifies women, then no other community is allowed to broadcast women-objectifying media at it. If a community wants to live an anarcho-primitivist lifestyle, nobody else is allowed to import TVs. Every community decides *exactly* how much informational contact it wants to have with the rest of the continent, and no one is allowed to force them to have more than that.

But the wizard and UniGov's most important task is to think of the children.

Imagine you're conservative Christians, and you're tired of this secular godless world, so you go off with your conservative Christian friends to found a conservative Christian community. You all pray together and stuff and are really happy. Then you have a daughter. Turns out she's atheist and lesbian. What now?

Well, it might be that your kid would be much happier at the lesbian separatist community the next island over. The *absolute minimum* the united government can do is enforce freedom of movement. That is, the *second* your daughter decides she doesn't want to be in Christiantopia anymore, she goes to a UniGov embassy nearby and asks for a ticket out, which they give her, free of charge. She gets airlifted to Lesbiantopia the next day. If *anyone* in Christiantopia tries to prevent her from reaching that embassy, or threatens her family if she leaves, or expresses the *slightest* amount of coercion to keep her around, UniGov burns their city and salts their field.

But this is not nearly enough to fully solve the child problem. A child who is abused may be too young to know that escape is an option, or may be brainwashed into thinking they are evil, or guilted into believing they are betraying their families to opt out. And although there is no perfect, elegant solution here, the practical solution is that UniGov enforces some pretty strict laws on child-rearing, and every child, no matter what other education they receive, also has to receive a class taught by a UniGov representative in which they learn about the other communities in the Archipelago, receive a basic non-brainwashed view of the world, and are given directions to their nearest UniGov representative who they can give their opt-out request to.

The list of communities they are informed about always starts with the capital, ruled by UniGov itself and considered an inoffensive, neutral option for people who don't want anywhere in particular. And it always ends with a reminder that if they can gather enough support, UniGov will provide them with a galleon to go out and found their own community in hitherto uninhabited lands.

There's one more problem UniGov has to deal with: malicious inter-community transfer. Suppose that there is some community which puts extreme effort into educating its children, an education which it supports through heavy taxation. New parents move to this community, reap the benefits, and then when their children grow up they move back to their previous community so they don't have to pay the taxes to educate anyone else. The communities themselves prevent some of this by

immigration restrictions – anyone who's clearly taking advantage of them isn't allowed in (except in the capital, which has an official committment to let in anyone who wants). But that still leaves the example of people maliciously leaving a high-tax community once they've got theirs. I imagine this is a big deal in Archipelago politics, but that in practice UniGov asks these people, even in their new homes, to pay higher tax rates to subsidize their old community. Or since that could be morally objectionable (imagine the lesbian separatist having to pay taxes to Christiantopia which oppressed her), maybe they pay the excess taxes to UniGov itself, just as a way of disincentivizing malicious movement.

Because there *are* UniGov taxes, and most people are happy to pay them. In my fantasy, UniGov isn't an enemy, where the Christians view it as this evil atheist conglomerate trying to steal their kids away from them and the capitalists view it as this evil socialist conglomerate trying to enforce high taxes. The Christians, the capitalists, and everyone else are extraordinarily *patriotic* about being part of the Archipelago, for its full name is the Archipelago of Civilized Communities, it is the standard-bearer of civilization against the barbaric outside world, and it is precisely the institution that allows them to maintain their distinctiveness in the face of what would otherwise be irresistable pressure to conform. Atheistopia is the enemy of Christiantopia, but only in the same way the Democratic Party is the enemy of the Republican Party – two groups within the same community who may have different ideas but who consider themselves part of the same broader whole, fundamentally allies under a banner of which both are proud.

IV.

Robert Nozick once proposed a similar idea as a libertarian utopia, and it's easy to see why. UniGov does very very little. Other than the part with children and the part with evening out taxation regimes, it just sits around preventing communities from using force against each other. That makes it very very easy for anyone who wants freedom to start a community that grants them the kind of freedom they want – or, more likely, to just start a community organized on purely libertarian principles. The United Government of Archipelago is the perfect minarchist night watchman state, and any additions you make over that are chosen by your own free will.

But other people could view the same plan as a conservative utopia. Conservativism, when it's not just Libertarianism Lite, is about building strong cohesive communities of relatively similar people united around common values. Archipelago is obviously built to make this as easy as possible, and it's hard to imagine that there wouldn't pop up a bunch of communities built around the idea of Decent Small-Town God-Fearing People where everyone has white picket fences and goes to the same church and nobody has to lock their doors at night (so basically Utah; I feel like this is one of the rare cases where the US' mostly-in-name-only Archipelagoness really asserts itself). People who didn't fit in could go to a Community Of People Who Don't Fit In and would have no need to nor right to complain, and no one would have to deal with Those Durned Bureaucrats In Washington telling them what to do.

But to me, this seems like a liberal utopia, even a leftist utopia, for three reasons.

The first reason is that it extends the basic principle of liberalism – solve differences of opinion by letting everyone do their own thing according to their own values, then celebrate the diversity this produces. I like homosexuality, you don't, fine, I can be homosexual and you don't have to, and having both gay and straight people living side by side enriches society. This just takes the whole thing one meta-level up – I

want to live in a very sexually liberated community, you want to live in a community where sex is treated purely as a sacred act for the purpose of procreation, fine, I can live in the community I want and you can live in the community you want, and having both sexually-liberated and sexually-pure communities living side by side enriches society. It is pretty much saying that the solution to any perceived problems of liberalism is *much more liberalism*.

The second reason is quite similar to the conservative reason. A lot of liberals have some pretty strong demands about the sorts of things they want society to do. I was recently talking to Ozy about a group who believe that society billing thin people is fatphobic, and that everyone needs to admit obese people can be just as attractive and date more of them, and that anyone who preferentially dates thinner people is Problematic. They also want people to stop talking about nutrition and exercise publicly. I sympathize with these people, especially having recently read a study showing that obese people are much happier when surrounded by other obese, rather than skinny people. But realistically, their movement will fail, and even philosophically, I'm not sure how to determine if they have the right to demand what they are demanding or what that question means. Their best bet is to found a community on these kinds of principles and only invite people who already share their preferences and aesthetics going in.

The third reason is the reason I specifically draw leftism in here. Liberalism, and to a much greater degree leftism, are marked by the emphasis they place on oppression. They're particularly marked by an emphasis on oppression being a really hard problem, and one that is structurally inherent to a certain society. They are marked by a moderate amount of despair that this oppression can ever be rooted out.

And I think a pretty strong response to this is making sure everyone is able to say "Hey, you better not oppress us, because if you do, we can pack up and go somewhere else."

Like if you want to protest that this is unfair, that people shouldn't be forced to leave their homes because of oppression, fine, fair enough. But given that oppression *is* going on, and you haven't been able to fix it, giving people the *choice* to get away from it seems like a pretty big win. I am reminded of the many Jews who moved from Eastern Europe to America, the many blacks who moved from the southern US to the northern US or Canada, and the many gays who make it out of extremely homophobic areas to friendlier large cities. One could even make a metaphor, I think rightly, to telling battered women that they are allowed to leave their husbands, telling them they're not forced to stay in a relationship that they consider abusive, and making sure that there are shelters available to receive them.

If any person who feels oppressed can leave whenever they like, to the point of being provided a free plane ticket by the government, how long can oppression go on before the oppressors give up and say "Yeah, guess we need someone to work at these factories now that all our workers have gone to the communally-owned factory down the road, we should probably at least let people unionize or something so they will tolerate us"?

A commenter in the latest Asch thread mentioned an interesting quote by Frederick Douglass:

> The American people have always been anxious to know what they shall do with us [black people]. I have had but one answer from the beginning. Do nothing with

> us! Your doing with us has already played the mischief with us. Do nothing with us!

It sounds like, if Frederick Douglass had the opportunity to go to some other community, or even found a black ex-slave community, no racists allowed, he probably would have taken it [edit: or not, or had strict conditions]. If the people in slavery during his own time period had had the chance to leave their plantations for that community, I bet they would have taken it too. And if you believe there are still people today whose relationship with society are similar in kind, if not in degree, to that of a plantation slave, you should be pretty enthusiastic about the ability of exit rights and free association to disrupt those oppressive relationships.

V.

We lack Archipelago's big advantage – a vast frontier of unsettled land.

Which is not to say that people don't form communes. They do. Some people even have really clever ideas along these lines, like the seasteaders. But the United States isn't going to become Archipelago any time soon.

There's another problem too, which I describe in my Anti-Reactionary FAQ. Discussing 'exit rights', I say:

> Exit rights are a great idea and of course having them is better than not having them. But I have yet to hear Reactionaries who cite them as a panacea explain in detail what exit rights we need beyond those we have already.
>
> The United States allows its citizens to leave the country by buying a relatively cheap passport and go anywhere that will take them in, with the exception of a few arch-enemies like Cuba – and those exceptions are laughably easy to evade. It allows them to hold dual citizenship with various foreign powers. It even allows them to renounce their American citizenship entirely and become sole citizens of any foreign power that will accept them.
>
> Few Americans take advantage of this opportunity in any but the most limited ways. When they do move abroad, it's usually for business or family reasons, rather than a rational decision to move to a different country with policies more to their liking. There are constant threats by dissatisfied Americans to move to Canada, and one in a thousand even carry through with them, but the general situation seems to be that America has a very large neighbor that speaks the same language, and has an equally developed economy, and has policies that many Americans prefer to their own country's, and isn't too hard to move to, and almost no one takes advantage of this opportunity. Nor do I see many people, even among the rich, moving to Singapore or Dubai.
>
> Heck, the US has fifty states. Moving from one to another is as easy as getting in a car, driving there, and renting a room, and although the federal government limits exactly how different their policies can be you better believe that there are very important differences in areas like taxes, business climate, education, crime, gun control, and many more. Yet aside from the fascinating but small-scale Free State Project there's little politically-motivated interstate movement, nor do states seem to have been motivated to converge on their policies or be less ideologically driven.
>
> What if we held an exit rights party, and nobody came?

> Even aside from the international problems of gaining citizenship, dealing with a language barrier, and adapting to a new culture, people are just rooted – property, friends, family, jobs. The end result is that the only people who can leave their countries behind are very poor refugees with nothing to lose, and very rich jet-setters. The former aren't very attractive customers, and the latter have all their money in tax shelters anyway.
>
> So although the idea of being able to choose your country like a savvy consumer appeals to me, just saying "exit rights!" isn't going to make it happen, and I haven't heard any more elaborate plans.

I guess I still feel that way. So although Archipelago is an interesting exercise in political science, a sort of pure case we can compare ourselves to, it doesn't look like a practical solution for real problems.

On the other hand, I do think it's worth becoming more Archipelagian on the margin rather than less so, and that there are good ways to do it.

One of the things that started this whole line of thought was an argument on Facebook about a very conservative Christian law school trying to open up in Canada. They had lots of rules like how their students couldn't have sex before marriage and stuff like that. The Canadian province they were in was trying to deny them accreditation, because conservative Christians are icky. I think the exact arguments being used were that it was homophobic, because the conservative Christians there would probably frown on married gays and therefore gays couldn't have sex at all. Therefore, the law school shouldn't be allowed to exist. There were other arguments of about this caliber, but they all seemed to boil down to "conservative Christians are icky".

This very much annoyed me. Yes, conservative Christians are icky. And they should be allowed to form completely voluntary communities of icky people that enforce icky cultural norms and an insular society promoting ickiness, just like everyone else. If non-conservative-Christians don't like what they're doing, they should *not go to that law school*. Instead they can go to one of the dozens of other law schools that conform to their own philosophies. And if gays want a law school even friendlier to them than the average Canadian law school, they should be allowed to create some law school that only accepts gays and bans homophobes and teaches lots of courses on gay marriage law all the time.

Another person on the Facebook thread complained that this line of arguments leads to being okay with white separatists. And so it does. Fine. I think white separatists have *exactly* the right position about where the sort of white people who want to be white separatists should be relative to everyone else – separate. I am not sure what you think you are gaining by demanding that white separatists live in communities with a lot of black people in them, but I bet the black people in those communities aren't thanking you. Why would they want a white separatist as a neighbor? Why should they have to have one?

If people want to go do their own thing in a way that harms no one else, you *let* them. That's the Archipelagian way.

(someone will protest that Archipelagian voluntary freedom of association or disassociation could, in cases of enough racial prejudice, lead to segregation, and that segregation didn't work. Indeed it didn't. But I feel like a version of segregation in which black people actually had the legally mandated right to get away from white

people and remain completely unmolested by them – and where a white-controlled government wasn't in charge of divvying up resources between white and black communities – would have worked a lot better than the segregation we actually had. The segregation we actually *had* was one in which white and black communities were separate until white people wanted something from black people, at which case they waltzed in and took it. If communities were actually totally separate, government and everything, by definition it would be impossible for one to oppress the other. The black community might start with less, but that could be solved by some kind of reparations. The Archipelagian way of dealing with this issue would be for white separatists to have separate white communities, black separatists to have separate black communities, integrationists to have integrated communities, resdistributive taxation from wealthier communities going into less wealthy ones, and a strong central government ruthlessly enforcing laws against any community trying to hurt another. I don't think there's a single black person in the segregation-era South who wouldn't have taken that deal, and any black person who thinks the effect of whites on their community today is net negative should be pretty interested as well.)

This is one reason I find people who hate seasteads so distasteful. I mean, here's what Reuters has to say about seasteading:

> Fringe movements, of course, rarely cast themselves as obviously fringe. Racist, anti-civil rights forces cloaked themselves in the benign language of "state's rights". Anti-gay religious entities adopted the glossy, positive imagery of "family values". Similarly, though many Libertarians embrace a pseudo-patriotic apple pie nostalgia, behind this façade is a very un-American, sinister vision.
>
> Sure, most libertarians may not want to do away entirely with the idea of government or, for that matter, government-protected rights and civil liberties. But many do — and ironically vie for political power in a nation they ultimately want to destroy. Even the right-wing pundit Ann Coulter mocked the paradox of Libertarian candidates: "Get rid of government — but first, make me president!" Libertarians sowed the seeds of anti-government discontent, which is on the rise, and now want to harvest that discontent for a very radical, anti-America agenda. The image of libertarians living off-shore in their lawless private nation-states is just a postcard of the future they hope to build on land.
>
> Strangely, the libertarian agenda has largely escaped scrutiny, at least compared to that of social conservatives. The fact that the political class is locked in debate about whether Michele Bachmann or Rick Perry is more socially conservative only creates a veneer of mainstream legitimacy for the likes of Ron Paul, whose libertarianism may be even more extreme and dangerously un-patriotic. With any luck America will recognize anti-government extremism for what it is — before libertarians throw America overboard and render us all castaways.

Keep in mind this is because *some people want to go off and do their own thing in the middle of the ocean far away from everyone else without bothering anyone*. And the newspapers are trying to whip up a panic about "throwing America overboard".

So one way we could become more Archipelagian is just *trying not to yell at people who are trying to go off and doing their own thing quietly with a group of voluntarily consenting friends*.

But I think a better candidate for how to build a more Archipelagian world is to encourage the fracture of society into subcultures.

Like, transsexuals may not be able to go to a transsexual island somewhere and build Transtopia where anyone who misgenders anyone else gets thrown into a volcano. But of the transsexuals I know, a lot of them have lots of transsexual friends, their cissexual friends are all up-to-date on trans issues and don't do a lot of misgendering, and they have great social networks where they share information about what businesses and doctors are or aren't trans-friendly. They can take advantage of trigger warnings to make sure they expose themselves to only the sources that fit the values of their community, the information that would get broadcast if it was a normal community that could impose media norms. As Internet interaction starts to replace real-life interaction (and I think for a lot of people the majority of their social life is already on the Internet, and for some the majority of their economic life is as well) it becomes increasingly easy to limit yourself to transsexual-friendly spaces that keep bad people away.

The rationalist community is another good example. If I wanted, I could move to the Bay Area tomorrow and never have more than a tiny amount of contact with non-rationalists again. I could have rationalist roommates, live in a rationalist group house, try to date only other rationalists, try to get a job with a rationalist nonprofit like CFAR or a rationalist company like Quixey, and never have to deal with the benighted and depressing non-rationalist world again. Even without moving to the Bay Area, it's been pretty easy for me to keep a lot of my social life, both on- and off- line, rationalist-focused, and I don't regret this at all.

I don't know if the future will be virtual reality. I expect the post-singularity future will include something like VR, although that might be like describing teleportation as "basically a sort of pack animal". But how much the immediate pre-singularity world will make use of virtual reality, I don't know.

But I bet if it doesn't, it will be because virtual reality has been circumvented by things like social networks, bitcoin, and Mechanical Turk, which make it possible to do most of your interaction through the Internet even though you're not literally plugged into it.

And that seems to me like a pretty good start in creating an Archipelago. I already hang out with various Finns and Brits and Aussies a lot more closely than I do my next-door neighbors, and if we start using litecoin and someone else starts using dogecoin then I'll be more economically connected to them too. The degree to which I encounter certain objectifying or unvirtuous or triggering media already depends more on the moderation policies of Less Wrong and Slate Star Codex and who I block from my Facebook feed, than it does any laws about censorship of US media.

At what point are national governments rendered mostly irrelevant compared to the norms and rules of the groups of which we are voluntary members?

I don't know, but I kind of look forward to finding out. It seems like a great way to start searching for utopia, or at least getting some people away from their metaphorical abusive-husbands.

And the other thing is that I have pretty strong opinions on which communities are better than others. Some communities were founded by toxic people for ganging up with other toxic people to celebrate and magnify their toxicity, and these (surprise, surprise) tend to be toxic. Others were formed by very careful, easily-harmed people trying to exclude everyone who could harm them, and these tend to be pretty safe albeit sometimes overbearing. Other people hit some kind of sweet spot that makes

friendly people want to come in and angry people want to stay out, or just do a really good job choosing friends.

But I think the end result is that the closer you come to true freedom of association, the closer you get to a world where everyone is a member of more or less the community they deserve. That would be a pretty unprecedented bit of progress.

Meditations On Moloch

[Content note: Visions! omens! hallucinations! miracles! ecstasies! dreams! adorations! illuminations! religions!]

I.

Allan Ginsberg's famous poem, *Moloch*:

> What sphinx of cement and aluminum bashed open their skulls and ate up their brains and imagination?
>
> Moloch! Solitude! Filth! Ugliness! Ashcans and unobtainable dollars! Children screaming under the stairways! Boys sobbing in armies! Old men weeping in the parks!
>
> Moloch! Moloch! Nightmare of Moloch! Moloch the loveless! Mental Moloch! Moloch the heavy judger of men!
>
> Moloch the incomprehensible prison! Moloch the crossbone soulless jailhouse and Congress of sorrows! Moloch whose buildings are judgment! Moloch the vast stone of war! Moloch the stunned governments!
>
> Moloch whose mind is pure machinery! Moloch whose blood is running money! Moloch whose fingers are ten armies! Moloch whose breast is a cannibal dynamo! Moloch whose ear is a smoking tomb!
>
> Moloch whose eyes are a thousand blind windows! Moloch whose skyscrapers stand in the long streets like endless Jehovahs! Moloch whose factories dream and croak in the fog! Moloch whose smoke-stacks and antennae crown the cities!
>
> Moloch whose love is endless oil and stone! Moloch whose soul is electricity and banks! Moloch whose poverty is the specter of genius! Moloch whose fate is a cloud of sexless hydrogen! Moloch whose name is the Mind!
>
> Moloch in whom I sit lonely! Moloch in whom I dream Angels! Crazy in Moloch! Cocksucker in Moloch! Lacklove and manless in Moloch!
>
> Moloch who entered my soul early! Moloch in whom I am a consciousness without a body! Moloch who frightened me out of my natural ecstasy! Moloch whom I abandon! Wake up in Moloch! Light streaming out of the sky!
>
> Moloch! Moloch! Robot apartments! invisible suburbs! skeleton treasuries! blind capitals! demonic industries! spectral nations! invincible madhouses! granite cocks! monstrous bombs!
>
> They broke their backs lifting Moloch to Heaven! Pavements, trees, radios, tons! lifting the city to Heaven which exists and is everywhere about us!
>
> Visions! omens! hallucinations! miracles! ecstasies! gone down the American river!
>
> Dreams! adorations! illuminations! religions! the whole boatload of sensitive bullshit!

> Breakthroughs! over the river! flips and crucifixions! gone down the flood! Highs! Epiphanies! Despairs! Ten years' animal screams and suicides! Minds! New loves! Mad generation! down on the rocks of Time!
>
> Real holy laughter in the river! They saw it all! the wild eyes! the holy yells! They bade farewell! They jumped off the roof! to solitude! waving! carrying flowers! Down to the river! into the street!

What's always impressed me about this poem is its conception of civilization as an individual entity. You can almost see him, with his fingers of armies and his skyscraper-window eyes.

A lot of the commentators say Moloch represents capitalism. This is definitely a piece of it, even a big piece. But it doesn't quite fit. Capitalism, whose fate is a cloud of sexless hydrogen? Capitalism in whom I am a consciousness without a body? Capitalism, therefore granite cocks?

Moloch is introduced as the answer to a question – C. S. Lewis' question in Hierarchy Of Philosophers – *what does it*? Earth could be fair, and all men glad and wise. Instead we have prisons, smokestacks, asylums. What sphinx of cement and aluminum breaks open their skulls and eats up their imagination?

And Ginsberg answers: *Moloch does it.*

There's a passage in the *Principia Discordia* where Malaclypse complains to the Goddess about the evils of human society. "Everyone is hurting each other, the planet is rampant with injustices, whole societies plunder groups of their own people, mothers imprison sons, children perish while brothers war."

The Goddess answers: "What is the matter with that, if it's what you want to do?"

Malaclypse: "But nobody wants it! Everybody hates it!"

Goddess: "Oh. Well, then stop."

The implicit question is – if everyone hates the current system, who perpetuates it? And Ginsberg answers: "Moloch". It's powerful not because it's correct – nobody literally thinks an ancient Carthaginian demon causes everything – but because thinking of the system as an agent throws into relief the degree to which the system *isn't* an agent.

Bostrom makes an offhanded reference of the possibility of a dictatorless dystopia, one that every single citizen including the leadership hates but which nevertheless endures unconquered. It's easy enough to imagine such a state. Imagine a country with two rules: first, every person must spend eight hours a day giving themselves strong electric shocks. Second, if anyone fails to follow a rule (including this one), or speaks out against it, or fails to enforce it, all citizens must unite to kill that person. Suppose these rules were well-enough established by tradition that everyone expected them to be enforced.

So you shock yourself for eight hours a day, because you know if you don't everyone else will kill you, because if they don't, everyone else will kill *them*, and so on. Every single citizen hates the system, but for lack of a good coordination mechanism it endures. From a god's-eye-view, we can optimize the system to "everyone agrees to

stop doing this at once", but no one within the system is able to effect the transition without great risk to themselves.

And okay, this example is kind of contrived. So let's run through – let's say ten – real world examples of similar multipolar traps to really hammer in how important this is.

1. The Prisoner's Dilemma, as played by two very dumb libertarians who keep ending up on defect-defect. There's a much better outcome available if they could figure out the coordination, but coordination is *hard*. From a god's-eye-view, we can agree that cooperate-cooperate is a better outcome than defect-defect, but neither prisoner within the system can make it happen.

2. Dollar auctions. I wrote about this and even more convoluted versions of the same principle in Game Theory As A Dark Art. Using some weird auction rules, you can take advantage of poor coordination to make someone pay $10 for a one dollar bill. From a god's-eye-view, clearly people should not pay $10 for a on-er. From within the system, each individual step taken might be rational.

(Ashcans and unobtainable dollars!)

3. The fish farming story from my Non-Libertarian FAQ 2.0:

> As a thought experiment, let's consider aquaculture (fish farming) in a lake. Imagine a lake with a thousand identical fish farms owned by a thousand competing companies. Each fish farm earns a profit of $1000/month. For a while, all is well.
>
> But each fish farm produces waste, which fouls the water in the lake. Let's say each fish farm produces enough pollution to lower productivity in the lake by $1/month.
>
> A thousand fish farms produce enough waste to lower productivity by $1000/month, meaning none of the fish farms are making any money. Capitalism to the rescue: someone invents a complex filtering system that removes waste products. It costs $300/month to operate. All fish farms voluntarily install it, the pollution ends, and the fish farms are now making a profit of $700/month – still a respectable sum.
>
> But one farmer (let's call him Steve) gets tired of spending the money to operate his filter. Now one fish farm worth of waste is polluting the lake, lowering productivity by $1. Steve earns $999 profit, and everyone else earns $699 profit.
>
> Everyone else sees Steve is much more profitable than they are, because he's not spending the maintenance costs on his filter. They disconnect their filters too.
>
> Once four hundred people disconnect their filters, Steve is earning $600/month – less than he would be if he and everyone else had kept their filters on! And the poor virtuous filter users are only making $300. Steve goes around to everyone, saying "Wait! We all need to make a voluntary pact to use filters! Otherwise, everyone's productivity goes down."
>
> Everyone agrees with him, and they all sign the Filter Pact, except one person who is sort of a jerk. Let's call him Mike. Now everyone is back using filters again, except Mike. Mike earns $999/month, and everyone else earns $699/month.

> Slowly, people start thinking they too should be getting big bucks like Mike, and disconnect their filter for $300 extra profit…
>
> A self-interested person never has any incentive to use a filter. A self-interested person has some incentive to sign a pact to make everyone use a filter, but in many cases has a stronger incentive to wait for everyone else to sign such a pact but opt out himself. This can lead to an undesirable equilibrium in which no one will sign such a pact.

The more I think about it, the more I feel like this is the core of my objection to libertarianism, and that Non-Libertarian FAQ 3.0 will just be this one example copy-pasted two hundred times. From a god's-eye-view, we can say that polluting the lake leads to bad consequences. From within the system, no individual can prevent the lake from being polluted, and buying a filter might not be such a good idea.

4. The Malthusian trap, at least at its extremely pure theoretical limits. Suppose you are one of the first rats introduced onto a pristine island. It is full of yummy plants and you live an idyllic life lounging about, eating, and composing great works of art (you're one of those rats from *The Rats of NIMH*).

You live a long life, mate, and have a dozen children. All of them have a dozen children, and so on. In a couple generations, the island has ten thousand rats and has reached its carrying capacity. Now there's not enough food and space to go around, and a certain percent of each new generation dies in order to keep the population steady at ten thousand.

A certain sect of rats abandons art in order to devote more of their time to scrounging for survival. Each generation, a bit less of this sect dies than members of the mainstream, until after a while, no rat composes any art at all, and any sect of rats who try to bring it back will go extinct within a few generations.

In fact, it's not just art. Any sect at all that is leaner, meaner, and more survivalist than the mainstream will eventually take over. If one sect of rats altruistically decides to limit its offspring to two per couple in order to decrease overpopulation, that sect will die out, swarmed out of existence by its more numerous enemies. If one sect of rats starts practicing cannibalism, and finds it gives them an advantage over their fellows, it will eventually take over and reach fixation.

If some rat scientists predict that depletion of the island's nut stores is accelerating at a dangerous rate and they will soon be exhausted completely, a few sects of rats might try to limit their nut consumption to a sustainable level. Those rats will be outcompeted by their more selfish cousins. Eventually the nuts will be exhausted, most of the rats will die off, and the cycle will begin again. Any sect of rats advocating some action to stop the cycle will be outcompeted by their cousins for whom advocating *anything* is a waste of time that could be used to compete and consume.

For a bunch of reasons evolution is not quite as Malthusian as the ideal case, but it provides the prototype example we can apply to other things to see the underlying mechanism. From a god's-eye-view, it's easy to say the rats should maintain a comfortably low population. From within the system, each individual rat will follow its genetic imperative and the island will end up in an endless boom-bust cycle.

5. Capitalism. Imagine a capitalist in a cutthroat industry. He employs workers in a sweatshop to sew garments, which he sells at minimal profit. Maybe he would like to pay his workers more, or give them nicer working conditions. But he can't, because

that would raise the price of his products and he would be outcompeted by his cheaper rivals and go bankrupt. Maybe many of his rivals are nice people who would like to pay their workers more, but unless they have some kind of ironclad guarantee that none of them are going to defect by undercutting their prices they can't do it.

Like the rats, who gradually lose all values except sheer competition, so companies in an economic environment of *sufficiently intense competition* are forced to abandon all values except optimizing-for-profit or else be outcompeted by companies that optimized for profit better and so can sell the same service at a lower price.

(I'm not really sure how widely people appreciate the value of analogizing capitalism to evolution. Fit companies – defined as those that make the customer want to buy from them – survive, expand, and inspire future efforts, and unfit companies – defined as those no one wants to buy from – go bankrupt and die out along with their company DNA. The reasons Nature is red and tooth and claw are the same reasons the market is ruthless and exploitative)

From a god's-eye-view, we can contrive a friendly industry where every company pays its workers a living wage. From within the system, there's no way to enact it.

(Moloch whose love is endless oil and stone! Moloch whose blood is running money!)

6. The Two-Income Trap, as recently discussed on this blog. It theorized that sufficiently intense competition for suburban houses in good school districts meant that people had to throw away lots of other values – time at home with their children, financial security – to optimize for house-buying-ability or else be consigned to the ghetto.

From a god's-eye-view, if everyone agrees not to take on a second job to help win their competition for nice houses, then everyone will get exactly as nice a house as they did before, but only have to work one job. From within the system, absent a government literally willing to ban second jobs, everyone who doesn't get one will be left behind.

(Robot apartments! Invisible suburbs!)

7. Agriculture. Jared Diamond calls it the worst mistake in human history. Whether or not it was a mistake, it wasn't an *accident* – agricultural civilizations simply outcompeted nomadic ones, inevitable and irresistably. Classic Malthusian trap. Maybe hunting-gathering was more enjoyable, higher life expectancy, and more conducive to human flourishing – but in a state of *sufficiently intense competition* between peoples, in which agriculture with all its disease and oppression and pestilence was the more competitive option, everyone will end up agriculturalists or go the way of the Comanche Indians.

From a god's-eye-view, it's easy to see everyone should keep the more enjoyable option and stay hunter-gatherers. From within the system, each individual tribe only faces the choice of going agricultural or inevitably dying.

8. Arms races. Large countries can spend anywhere from 5% to 30% of their budget on defense. In the absence of war – a condition which has mostly held for the past fifty years – all this does is sap money away from infrastructure, health, education, or economic growth. But any country that fails to spend enough money on defense risks being invaded by a neighboring country that did. Therefore, almost all countries try to spend some money on defense.

From a god's-eye-view, the best solution is world peace and no country having an army at all. From within the system, no country can unilaterally enforce that, so their best option is to keep on throwing their money into missiles that lie in silos unused.

(Moloch the vast stone of war! Moloch whose fingers are ten armies!)

9. Cancer. The human body is supposed to be made up of cells living harmoniously and pooling their resources for the greater good of the organism. If a cell defects from this equilibrium by investing its resources into copying itself, it and its descendants will flourish, eventually outcompeting all the other cells and taking over the body – at which point it dies. Or the situation may repeat, with certain cancer cells defecting against the rest of the tumor, thus slowing down its growth and causing the tumor to stagnate.

From a god's-eye-view, the best solution is all cells cooperating so that they don't all die. From within the system, cancerous cells will proliferate and outcompete the other – so that only the existence of the immune system keeps the natural incentive to turn cancerous in check.

10. The "race to the bottom" describes a political situation where some jurisdictions lure businesses by promising lower taxes and fewer regulations. The end result is that either everyone optimizes for competitiveness – by having minimal tax rates and regulations – or they lose all of their business, revenue, and jobs to people who did (at which point they are pushed out and replaced by a government who will be more compliant).

But even though the last one has stolen the name, all these scenarios are in fact a race to the bottom. Once one agent learns how to become more competitive by sacrificing a common value, all its competitors must also sacrifice that value or be outcompeted and replaced by the less scrupulous. Therefore, the system is likely to end up with everyone once again equally competitive, but the sacrificed value is gone forever. From a god's-eye-view, the competitors know they will all be worse off if they defect, but from within the system, given insufficient coordination it's impossible to avoid.

Before we go on, there's a slightly different form of multi-agent trap worth investigating. In this one, the competition is kept at bay by some outside force – usually social stigma. As a result, there's not actually a race to the bottom – the system can continue functioning at a relatively high level – but it's impossible to optimize and resources are consistently thrown away for no reason. Lest you get exhausted before we even begin, I'll limit myself to four examples here.

11. Education. In my essay on reactionary philosophy, I talk about my frustration with education reform:

> People ask why we can't reform the education system. But right now students' incentive is to go to the most prestigious college they can get into so employers will hire them – whether or not they learn anything. Employers' incentive is to get students from the most prestigious college they can so that they can defend their decision to their boss if it goes wrong – whether or not the college provides value added. And colleges' incentive is to do whatever it takes to get more prestige, as measured in *US News and World Report* rankings – whether or not it helps students. Does this lead to huge waste and poor education? Yes. Could the Education God notice this and make some Education Decrees that lead to a vastly more efficient system? Easily! But since there's no Education God everybody is

> just going to follow their own incentives, which are only partly correlated with education or efficiency.

From a god's eye view, it's easy to say things like "Students should only go to college if they think they will get something out of it, and employers should hire applicants based on their competence and not on what college they went to". From within the system, everyone's already following their own incentives correctly, so unless the incentives change the system won't either.

12. Science. Same essay:

> The modern research community *knows* they aren't producing the best science they could be. There's lots of publication bias, statistics are done in a confusing and misleading way out of sheer inertia, and replications often happen very late or not at all. And sometimes someone will say something like "I can't believe people are too dumb to fix Science. All we would have to do is require early registration of studies to avoid publication bias, turn this new and powerful statistical technique into the new standard, and accord higher status to scientists who do replication experiments. It would be really simple and it would vastly increase scientific progress. I must just be smarter than all existing scientists, since I'm able to think of this and they aren't."
>
> And yeah. That would work for the Science God. He could just make a Science Decree that everyone has to use the right statistics, and make another Science Decree that everyone must accord replications higher status.
>
> But things that work from a god's-eye view don't work from within the system. No individual scientist has an incentive to unilaterally switch to the new statistical technique for her own research, since it would make her research less likely to produce earth-shattering results and since it would just confuse all the other scientists. They just have an incentive to want everybody else to do it, at which point they would follow along. And no individual journal has an incentive to unilaterally switch to early registration and publishing negative results, since it would just mean their results are less interesting than that other journal who only publishes ground-breaking discoveries. From within the system, everyone is following their own incentives and will continue to do so.

13. Government corruption. I don't know of anyone who really thinks, in a principled way, that corporate welfare is a good idea. But the government still manages to spend somewhere around (depending on how you calculate it) $100 billion dollars a year on it – which for example is three times the amount they spend on health care for the needy. Everyone familiar with the problem has come up with the same easy solution: stop giving so much corporate welfare. Why doesn't it happen?

Government are competing against one another to get elected or promoted. And suppose part of optimizing for electability is optimizing campaign donations from corporations – or maybe it isn't, but officials *think* it is. Officials who try to mess with corporate welfare may lose the support of corporations and be outcompeted by officials who promise to keep it intact.

So although from a god's-eye-view everyone knows that eliminating corporate welfare is the best solution, each individual official's personal incentives push her to maintain it.

14. Congress. Only 9% of Americans like it, suggesting a lower approval rating than cockroaches, head lice, or traffic jams. However, 62% of people who know who their own Congressional representative is approve of them. In theory, it should be *really hard* to have a democratically elected body that maintains a 9% approval rating for more than one election cycle. In practice, every representative's incentive is to appeal to his or her constituency while throwing the rest of the country under the bus – something at which they apparently succeed.

From a god's-eye-view, every Congressperson ought to think only of the good of the nation. From within the system, you do what gets you elected.

II.

A basic principle unites all of the multipolar traps above. In some competition optimizing for X, the opportunity arises to throw some other value under the bus for improved X. Those who take it prosper. Those who don't take it die out. Eventually, everyone's relative status is about the same as before, but everyone's absolute status is worse than before. The process continues until all other values that can be traded off have been – in other words, until human ingenuity cannot possibly figure out a way to make things any worse.

In a sufficiently intense competition (1-10), everyone who doesn't throw all their values under the bus dies out – think of the poor rats who wouldn't stop making art. This is the infamous Malthusian trap, where everyone is reduced to "subsistence".

In an insufficiently intense competition (11-14), all we see is a perverse failure to optimize – consider the journals which can't switch to more reliable science, or the legislators who can't get their act together and eliminate corporate welfare. It may not reduce people to subsistence, but there is a weird sense in which it takes away their free will.

Every two-bit author and philosopher has to write their own utopia. Most of them are legitimately pretty nice. In fact, it's a pretty good bet that two utopias that are polar opposites both sound better than our own world.

It's kind of embarrassing that random nobodies can think up states of affairs better than the one we actually live in. And in fact most of them can't. A lot of utopias sweep the hard problems under the rug, or would fall apart in ten minutes if actually implemented.

But let me suggest a couple of "utopias" that don't have this problem.

– The utopia where instead of the government paying lots of corporate welfare, the government *doesn't* pay lots of corporate welfare.

– The utopia where every country's military is 50% smaller than it is today, and the savings go into infrastructure spending.

– The utopia where all hospitals use the same electronic medical record system, or at least medical record systems that can talk to each other, so that doctors can look up what the doctor you saw last week in a different hospital decided instead of running all the same tests over again for $5000.

I don't think there are too many people who *oppose* any of these utopias. If they're not happening, it's not because people don't support them. It certainly isn't because

nobody's thought of them, since I just thought of them right now and I don't expect my "discovery" to be hailed as particularly novel or change the world.

Any human with above room temperature IQ can design a utopia. The reason our current system isn't a utopia is that *it wasn't designed by humans*. Just as you can look at an arid terrain and determine what shape a river will one day take by assuming water will obey gravity, so you can look at a civilization and determine what shape its institutions will one day take by assuming people will obey incentives.

But that means that just as the shapes of rivers are not designed for beauty or navigation, but rather an artifact of randomly determined terrain, so institutions will not be designed for prosperity or justice, but rather an artifact of randomly determined initial conditions.

Just as people can level terrain and build canals, so people can alter the incentive landscape in order to build better institutions. But they can only do so when they are incentivized to do so, which is not always. As a result, some pretty wild tributaries and rapids form in some very strange places.

I will now jump from boring game theory stuff to what might be the closest thing to a mystical experience I've ever had.

Like all good mystical experiences, it happened in Vegas. I was standing on top of one of their many tall buildings, looking down at the city below, all lit up in the dark. If you've never been to Vegas, it is *really* impressive. Skyscrapers and lights in every variety strange and beautiful all clustered together. And I had two thoughts, crystal clear:

It is glorious that we can create something like this.

It is shameful that we *did*.

Like, by what standard is building gigantic forty-story-high indoor replicas of Venice, Paris, Rome, Egypt, and Camelot side-by-side, filled with albino tigers, in the middle of the most inhospitable desert in North America, a remotely sane use of our civilization's limited resources?

And it occurred to me that maybe there is no philosophy on Earth that would endorse the existence of Las Vegas. Even Objectivism, which is usually my go-to philosophy for justifying the excesses of capitalism, at least grounds it in the belief that capitalism improves people's lives. Henry Ford was virtuous because he allowed lots of otherwise car-less people to obtain cars and so made them better off. What does Vegas do? Promise a bunch of shmucks free money and not give it to them.

Las Vegas doesn't exist because of some decision to hedonically optimize civilization, it exists because of a quirk in dopaminergic reward circuits, plus the microstructure of an uneven regulatory environment, plus Schelling points. A rational central planner with a god's-eye-view, contemplating these facts, might have thought "Hm, dopaminergic reward circuits have a quirk where certain tasks with slightly negative risk-benefit ratios get an emotional valence associated with slightly positive risk-benefit ratios, let's see if we can educate people to beware of that." People within the system, *following the incentives created by these facts*, think: "Let's build a forty-story-high indoor replica of ancient Rome full of albino tigers in the middle of the desert, and so become slightly richer than people who didn't!"

Just as the course of a river is latent in a terrain even before the first rain falls on it – so the existence of Caesar's Palace was latent in neurobiology, economics, and regulatory regimes even before it existed. The entrepreneur who built it was just filling in the ghostly lines with real concrete.

So we have all this amazing technological and cognitive energy, the brilliance of the human species, wasted on reciting the lines written by poorly evolved cellular receptors and blind economics, like gods being ordered around by a moron.

Some people have mystical experiences and see God. There in Las Vegas, I *saw* Moloch.

(Moloch, whose mind is pure machinery! Moloch, whose blood is running money!

Moloch whose soul is electricity and banks! Moloch, whose skyscrapers stand in the long streets like endless Jehovahs!

Moloch! Moloch! Robot apartments! Invisible suburbs! Skeleton treasuries! Blind capitals! Demonic industries! Spectral nations!)

…granite cocks!

III.

The Apocrypha Discordia says:

> Time flows like a river. Which is to say, downhill. We can tell this because everything is going downhill rapidly. It would seem prudent to be somewhere else when we reach the sea.

Let's take this random gag 100% literally and see where it leads us.

We just analogized the flow of incentives to the flow of a river. The downhill trajectory is appropriate: the traps happen when you find an opportunity to trade off a useful value for greater competitiveness. Once everyone has it, the greater competitiveness brings you no joy – but the value is lost forever. Therefore, each step of the Poor Coordination Polka makes your life worse.

But not only have we not yet reached the sea, but we also seem to move *uphill* surprisingly often. Why do things not degenerate more and more until we are back at subsistence level? I can think of three bad reasons – excess resources, physical limitations, and utility maximization – plus one good reason – coordination.

1. Excess resources. The ocean depths are a horrible place with little light, few resources, and various horrible organisms dedicated to eating or parasitizing one another. But every so often, a whale carcass falls to the bottom of the sea. More food than the organisms that find it could ever possibly want. There's a brief period of miraculous plenty, while the couple of creatures that first encounter the whale feed like kings. Eventually more animals discover the carcass, the faster-breeding animals in the carcass multiply, the whale is gradually consumed, and everyone sighs and goes back to living in a Malthusian death-trap.

(Slate Star Codex: Your source for macabre whale metaphors since June 2014)

It's as if a group of those rats who had abandoned art and turned to cannibalism suddenly was blown away to a new empty island with a much higher carrying capacity, where they would once again have the breathing room to live in peace and create artistic masterpieces.

This is an age of whalefall, an age of excess carrying capacity, an age when we suddenly find ourselves with a thousand-mile head start on Malthus. As Hanson puts it, this is the dream time.

As long as resources aren't scarce enough to lock us in a war of all against all, we can do silly non-optimal things – like art and music and philosophy and love – and not be outcompeted by merciless killing machines most of the time.

2. Physical limitations. Imagine a profit-maximizing slavemaster who decided to cut costs by not feeding his slaves or letting them sleep. He would soon find that his slaves' productivity dropped off drastically, and that no amount of whipping them could restore it. Eventually after testing numerous strategies, he might find his slaves got the most work done when they were well-fed and well-rested and had at least a little bit of time to relax. Not because the slaves were voluntarily withholding their labor – we assume the fear of punishment is enough to make them work as hard as they can – but because the body has certain physical limitations that limit how mean you can get away with being. Thus, the "race to the bottom" stops somewhere short of the actual ethical bottom, when the physical limits are run into.

John Moes, a historian of slavery, goes further and writes about how the slavery we are most familiar with – that of the antebellum South – is a historical aberration and probably economically inefficient. In most past forms of slavery – especially those of the ancient world – it was common for slaves to be paid wages, treated well, and often given their freedom.

He argues that this was the result of rational economic calculation. You can incentivize slaves through the carrot or the stick, and the stick isn't very good. You can't watch slaves all the time, and it's really hard to tell whether a slave is slacking off or not (or even whether, given a little more whipping, he might be able to work even harder). If you want your slaves to do anything more complicated than pick cotton, you run into some serious monitoring problems – how do you profit from an enslaved philosopher? Whip him really hard until he elucidates a theory of The Good that you can sell books about?

The ancient solution to the problem – perhaps an early inspiration to Fnargl – was to tell the slave to go do whatever he wanted and found most profitable, then split the profits with him. Sometimes the slave would work a job at your workshop and you would pay him wages based on how well he did. Other times the slave would go off and make his way in the world and send you some of what he earned. Still other times, you would set a price for the slave's freedom, and the slave would go and work and eventually come up with the mone and free himself.

Moes goes even further and says that these systems were so profitable that there were constant smouldering attempts to try this sort of thing in the American South. The reason they stuck with the whips-and-chains method owed less to economic considerations and more to racist government officials cracking down on lucrative but not-exactly-white-supremacy-promoting attempts to free slaves and have them go into business.

So in this case, a race to the bottom where competing plantations become crueler and crueler to their slaves in order to maximize competitiveness is halted by the physical limitation of cruelty not helping after a certain point.

Or to give another example, one of the reasons we're not currently in a Malthusian population explosion right now is that women can only have one baby per nine months. If those weird religious sects that demand their members have as many babies as possible could copy-paste themselves, we would be in *really* bad shape. As it is they can only do a small amount of damage per generation.

3. Utility maximization. We've been thinking in terms of preserving values versus winning competitions, and expecting optimizing for the latter to destroy the former.

But many of the most important competitions / optimization processes in modern civilization are optimizing for human values. You win at capitalism partly by satisfying customers' values. You win at democracy partly by satisfying voters' values.

Suppose there's a coffee plantation somewhere in Ethiopia that employs Ethiopians to grow coffee beans that get sold to the United States. Maybe it's locked in a life-and-death struggle with other coffee plantations and want to throw as many values under the bus as it can to pick up a slight advantage.

But it can't sacrifice quality of coffee produced too much, or else the Americans won't buy it. And it can't sacrifice wages or working conditions too much, or else the Ethiopians won't work there. And in fact, part of its competition-optimization process is finding the best ways to attract workers and customers that it can, as long as it doesn't cost them too much money. So this is very promising.

But it's important to remember exactly how fragile this beneficial equilibrium is.

Suppose the coffee plantations discover a toxic pesticide that will increase their yield but make their customers sick. But their customers don't know about the pesticide, and the government hasn't caught up to regulating it yet. Now there's a tiny uncoupling between "selling to Americans" and "satisfying Americans' values", and so of course Americans' values get thrown under the bus.

Or suppose that there's a baby boom in Ethiopia and suddenly there are five workers competing for each job. Now the company can afford to lower wages and implement cruel working conditions down to whatever the physical limits are. As soon as there's an uncoupling between "getting Ethiopians to work here" and "satisfying Ethiopian values", it doesn't look too good for Ethiopian values either.

Or suppose someone invents a robot that can pick coffee better and cheaper than a human. The company fires all its laborers and throws them onto the street to die. As soon as the utility of the Ethiopians is no longer necessary for profit, all pressure to maintain it disappears.

Or suppose that there is some important value that is neither a value of the employees or the customers. Maybe the coffee plantations are on the habitat of a rare tropical bird that environmentalist groups want to protect. Maybe they're on the ancestral burial ground of a tribe different from the one the plantation is employing, and they want it respected in some way. Maybe coffee growing contributes to global warming somehow. As long as it's not a value that will prevent the average American from buying from them or the average Ethiopian from working for them, under the bus it goes.

I know that “capitalists sometimes do bad things” isn’t exactly an original talking point. But I do want to stress how it’s not equivalent to “capitalists are greedy”. I mean, sometimes they *are* greedy. But other times they’re just in a sufficiently intense competition where anyone who doesn’t do it will be outcompeted and replaced by people who do. Business practices are set by Moloch, no one else has any choice in the matter.

(from my very little knowledge of Marx, he understands this very very well and people who summarize him as “capitalists are greedy” are doing him a disservice)

And as well understood as the capitalist example is, I think it is less well appreciated that democracy has the same problems. Yes, in theory it’s optimizing for voter happiness which correlates with good policymaking. But as soon as there’s the slightest disconnect between good policymaking and electability, good policymaking *has to* get thrown under the bus.

For example, ever-increasing prison terms are unfair to inmates and unfair to the society that has to pay for them. Politicans are unwilling to do anything about them because they don’t want to look “soft on crime”, and if a single inmate whom they helped release ever does anything bad (and statistically one of them will have to) it will be all over the airwaves as “Convict released by Congressman’s policies kills family of five, how can the Congressman even sleep at night let alone claim he deserves reelection?”. So even if decreasing prison populations would be good policy – and it is – it will be very difficult to implement.

(Moloch the incomprehensible prison! Moloch the crossbone soulless jailhouse and Congress of sorrows! Moloch whose buildings are judgment! Moloch the stunned governments!)

Turning “satisfying customers” and “satisfying citizens” into the *outputs* of optimization processes was one of civilization’s greatest advances and the reason why capitalist democracies have so outperformed other systems. But if we have bound Moloch as our servant, the bonds are not very strong, and we sometimes find that the tasks he has done for us move to his advantage rather than ours.

4. Coordination.

The opposite of a trap is a garden.

Things are easy to solve from a god’s-eye-view, so if everyone comes together into a superorganism, that superorganism can solve problems with ease and finesse. An intense competition between agents has turned into a garden, with a single gardener dictating where everything should go and removing elements that do not conform to the pattern.

As I pointed out in the Non-Libertarian FAQ, government can easily solve the pollution problem with fish farms. The best known solution to the Prisoners’ Dilemma is for the mob boss (playing the role of a governor) to threaten to shoot any prisoner who defects. The solution to companies polluting and harming workers is government regulations against such. Governments solve arm races *within* a country by maintaining a monopoly on the use of force, and it’s easy to see that if a truly effective world government ever arose, international military buildups would end pretty quickly.

The two active ingredients of government are laws plus violence – or more abstractly agreements plus enforcement mechanism. Many other things besides governments share these two active ingredients and so are able to act as coordination mechanisms to avoid traps.

For example, since students are competing against each other (directly if classes are graded on a curve, but always indirectly for college admissions, jobs, et cetera) there is intense pressure for individual students to cheat. The teacher and school play the role of a government by having rules (for example, against cheating) and the ability to punish students who break them.

But the emergent social structure of the students themselves is also a sort of government. If students shun and distrust cheaters, then there are rules (don't cheat) and an enforcement mechanism (or else we will shun you).

Social codes, gentlemens' agreements, industrial guilds, criminal organizations, traditions, friendships, schools, corporations, and religions are all coordinating institutions that keep us out of traps by changing our incentives.

But these institutions not only incentivize others, but are incentivized themselves. These are large organizations made of lots of people who are competing for jobs, status, prestige, et cetera – there's no reason they should be immune to the same multipolar traps as everyone else, and indeed they aren't. Governments can in theory keep corporations, citizens, et cetera out of certain traps, but as we saw above there are many traps that governments themselves can fall into.

The United States tries to solve the problem by having multiple levels of government, unbreakable constutitional laws, checks and balances between different branches, and a couple of other hacks.

Saudi Arabia uses a different tactic. They just put one guy in charge of everything.

This is the much-maligned – I think unfairly – argument in favor of monarchy. A monarch is an unincentivized incentivizer. He *actually* has the god's-eye-view and is outside of and above every system. He has permanently won all competitions and is not competing for anything, and therefore he is perfectly free of Moloch and of the incentives that would otherwise channel his incentives into predetermined paths. Aside from a few very theoretical proposals like my Shining Garden, monarchy is the *only* system that does this.

But then instead of following a random incentive structure, we're following the whim of one guy. Caesar's Palace Hotel and Casino is a crazy waste of resources, but the actual Gaius Julius Caesar Augustus Germanicus wasn't exactly the perfect benevolent rational central planner either.

The libertarian-authoritarian axis on the Political Compass is a tradeoff between discoordination and tyranny. You can have everything perfectly coordinated by someone with a god's-eye-view – but then you risk Stalin. And you can be totally free of all central authority – but then you're stuck in every stupid multipolar trap Moloch can devise.

The libertarians make a convincing argument for the one side, and the monarchists for the other, but I expect that like most tradeoffs we just have to hold our noses and admit it's a really hard problem.

IV.

Let's go back to that Apocrypha Discordia quote:

> Time flows like a river. Which is to say, downhill. We can tell this because everything is going downhill rapidly. It would seem prudent to be somewhere else when we reach the sea.

What would it mean, in this situation, to reach the sea?

Multipolar traps – races to the bottom – threaten to destroy all human values. They are currently restrained by physical limitations, excess resources, utility maximization, and coordination.

The dimension along which this metaphorical river flows must be time, and the most important change in human civilization over time is the change in technology. So the relevant question is how technological changes will affect our tendency to fall into multipolar traps.

I described traps as when:

> …in some competition optimizing for X, the opportunity arises to throw some other value under the bus for improved X. Those who take it prosper. Those who don't take it die out. Eventually, everyone's relative status is about the same as before, but everyone's absolute status is worse than before. The process continues until all other values that can be traded off have been – in other words, until human ingenuity cannot possibly figure out a way to make things any worse.

That "the opportunity arises" phrase is looking pretty sinister. Technology is all about creating new opportunities.

Develop a new robot, and suddenly coffee plantations have "the opportunity" to automate their harvest and fire all the Ethiopian workers. Develop nuclear weapons, and suddenly countries are stuck in an arms race to have enough of them. Polluting the atmosphere to build products quicker wasn't a problem before they invented the steam engine.

The limit of multipolar traps as technology approaches infinity is "very bad".

Multipolar traps are currently restrained by physical limitations, excess resources, utility maximization, and coordination.

Physical limitations are most obviously conquered by increasing technology. The slavemaster's old conundrum – that slaves need to eat and sleep – succumbs to Soylent and modafinil. The problem of slaves running away succumbs to GPS. The problem of slaves being too stressed to do good work succumbs to Valium. None of these things are very good for the slaves.

(or just invent a robot that doesn't need food or sleep at all. What happens to the slaves after that is better left unsaid)

The other example of physical limits was one baby per nine months, and this was understating the case – it's really "one baby per nine months plus willingness to support and take care of a basically helpless and extremely demanding human being

for eighteen years". This puts a damper on the enthusiasm of even the most zealous religious sect's "go forth and multiply" dictum.

But as Bostrom puts it in *Superintelligence*:

> There are reasons, if we take a longer view and assume a state of unchanging technology and continued prosperity, to expect a return to the historically and ecologically normal condition of a world population that butts up against the limits of what our niche can support. If this seems counterintuitive in light of the negative relationship between wealth and fertility that we are currently observing on the global scale, we must remind ourselves that this modern age is a brief slice of history and very much an aberration. Human behavior has not yet adapted to contemporary conditions. Not only do we fail to take advantage of obvious ways to increase our inclusive fitness (such as by becoming sperm or egg donors) but we actively sabotage our fertility by using birth control. In the environment of evolutionary adaptedness, a healthy sex drive may have been enough to make an individual act in ways that maximized her reproductive potential; in the modern environment, however, there would be a huge selective advantage to having a more direct desire for being the biological parent to the largest possible number of chilren. Such a desire is currently being selected for, as are other traits that increase our propensity to reproduce. Cultural adaptation, however, might steal a march on biological evolution. Some communities, such as those of the Hutterites or the adherents of the Quiverfull evangelical movement, have natalist cultures that encourage large families, and they are consequently undergoing rapid expansion…This longer-term outlook could be telescoped into a more imminent prospect by the intelligence explosion. Since software is copyable, a population of emulations or AIs could double rapidly – over the course of minutes rather than decades or centuries – soon exhausting all available hardware

As always when dealing with high-level transhumanists, "all available hardware" should be taken to include "the atoms that used to be part of your body".

The idea of biological *or* cultural evolution causing a mass population explosion is a philosophical toy at best. The idea of technology making it possible is both plausible and terrifying. Now we see that "physical limits" segues very naturally into "excess resources" – the ability to create new agents very quickly means that unless everyone can coordinate to ban doing this, the people who do will outcompete the people who don't until they have reached carrying capacity and everyone is stuck at subsistence level.

Excess resources, which until now have been a gift of technological progress, therefore switch and become a casualty of it at a sufficiently high tech level.

Utility maximization, always on shaky ground, also faces new threats. In the face of continuing debate about this point, I *continue* to think it obvious that robots will push humans out of work or at least drive down wages (which, in the existence of a minimum wage, pushes humans out of work).

Once a robot can do everything an IQ 80 human can do, only better and cheaper, there will be no reason to employ IQ 80 humans. Once a robot can do everything an IQ 120 human can do, only better and cheaper, there will be no reason to employ IQ 120 humans. Once a robot can do everything an IQ 180 human can do, only better and cheaper, there will be no reason to employ humans at all, in the unlikely scenario that there are any left by that point.

In the earlier stages of the process, capitalism becomes more and more uncoupled from its previous job as an optimizer for human values. Now most humans are totally locked out of the group whose values capitalism optimizes for. They have no value to contribute as workers – and since in the absence of a spectacular social safety net it's unclear how they would have much money – they have no value as customers either. Capitalism has passed them by. As the segment of humans who can be outcompeted by robots increases, capitalism passes by more and more people until eventually it locks out the human race entirely, once again in the vanishingly unlikely scenario that we are still around.

(there are some scenarios in which a few capitalists who own the robots may benefit here, but in either case the vast majority are out of luck)

Democracy is less obviously vulnerable, but it might be worth going back to Bostrom's paragraph about the Quiverfull movement. These are some really religious Christians who think that God wants them to have as many kids as possible, and who can end up with families of ten or more. Their articles explictly calculate that if they start at two percent of the population, but have on average eight children per generation when everyone else on average only has two, within three generations they'll make up half the population.

It's a clever strategy, but I can think of one thing that will save us: judging by how many ex-Quiverfull blogs I found when searching for those statistics, their retention rates even within a single generation are pretty grim. Their article admits that 80% of very religious children leave the church as adults (although of course they expect their own movement to do better). And this is not a symmetrical process – 80% of children who grow up in atheist families aren't becoming Quiverfull.

It looks a lot like even though they are outbreeding us, we are outmeme-ing them, and that gives us a decisive advantage.

But we should also be kind of scared of this process. Memes optimize for making people want to accept them and pass them on – so like capitalism and democracy, they're optimizing for a *proxy* of making us happy, but that proxy can easily get uncoupled from the original goal.

Chain letters, urban legends, propaganda, and viral marketing are all examples of memes that don't satisfy our explicit values (true and useful) but are sufficiently memetically virulent that they spread anyway.

I hope it's not too controversial here to say the same thing is true of religion. Religions, at their heart, are the most basic form of memetic replicator – "Believe this statement and repeat it to everyone you hear or else you will be eternally tortured".

The creationism "debate" and global warming "debate" and a host of similar "debates" in today's society suggest that memes that can propagate independent of their truth value has a pretty strong influence on the political process. Maybe these memes propagate because they appeal to people's prejudices, maybe because they're simple, maybe because they effectively mark an in-group and an out-group, or maybe for all sorts of different reasons.

The point is – imagine a country full of bioweapon labs, where people toil day and night to invent new infectious agents. The existence of these labs, and their right to throw whatever they develop in the water supply is protected by law. And the country is also linked by the world's most perfect mass transit system that every single person

uses every day, so that any new pathogen can spread to the entire country instantaneously. You'd expect things to start going bad for that city pretty quickly.

Well, we have about a zillion think tanks researching new and better forms of propaganda. And we have constitutionally protected freedom of speech. And we have the Internet. So we're kind of screwed.

(Moloch whose name is the Mind!)

There are a few people working on raising the sanity waterline, but not as many people as are working on new and exciting ways of confusing and converting people, cataloging and exploiting every single bias and heuristic and dirty rhetorical trick

So as technology (which I take to include knowledge of psychology, sociology, public relations, etc) tends to infinity, the power of truthiness relative to truth increases, and things don't look great for real grassroots democracy. The worst-case scenario is that the ruling party learns to produce infinite charisma on demand. If that doesn't sound so bad to you, remember what Hitler was able to do with an famously high level of charisma that was still less-than-infinite.

(alternate phrasing for Chomskyites: technology increases the efficiency of manufacturing consent in the same way it increases the efficiency of manufacturing everything else)

Coordination is what's left. And technology has the potential to seriously *improve* coordination efforts. People can use the Internet to get in touch with one another, launch political movements, and fracture off into subcommunities.

But coordination only works when you have 51% or more of the force on the side of the people doing the coordinating, and when you haven't come up with some brilliant trick to make coordination impossible.

The second one first. In the links post before last, I wrote:

> The latest development in the brave new post-Bitcoin world is crypto-equity. At this point I've gone from wanting to praise these inventors as bold libertarian heroes to wanting to drag them in front of a blackboard and making them write a hundred times "I WILL NOT CALL UP THAT WHICH I CANNOT PUT DOWN"

A couple people asked me what I meant, and I didn't have the background then to explain. Well, this post is the background. People are using the *contingent* stupidity of our current government to replace lots of human interaction with mechanisms that cannot be coordinated even in principle. I totally understand why all these things are good right now when most of what our government does is stupid and unnecessary. But there is going to come a time when – after one too many bioweapon or nanotech or nuclear incidents – we, as a civilization, are going to wish we hadn't established untraceable and unstoppable ways of selling products.

And if we ever get real live superintelligence, pretty much by definition it is going to have >51% of the power and all attempts at "coordination" with it will be useless.

So I agree with Robin Hanson: This is the dream time. This is a rare confluence of circumstances where the we are unusually safe from multipolar traps, and as such weird things like art and science and philosophy and love can flourish.

As technological advance increases, the rare confluence will come to an end. New opportunities to throw values under the bus for increased competitiveness will arise. New ways of copying agents to increase the population will soak up our excess resources and resurrect Malthus' unquiet spirit. Capitalism and democracy, previously our protectors, will figure out ways to route around their inconvenient dependence on human values. And our coordination power will not be nearly up to the task, assuming somthing much more powerful than all of us combined doesn't show up and crush our combined efforts with a wave of its paw.

Absent an extraordinary effort to divert it, the river reaches the sea in one of two places.

It can end in Eliezer Yudkowsky's nightmare of a superintelligence optimizing for some random thing (classically paper clips) because we weren't smart enough to channel its optimization efforts the right way. This is the ultimate trap, the trap that catches the universe. Everything except the one thing being maximized is destroyed utterly in pursuit of the single goal, including all the silly human values.

Or it can end in Robin Hanson's nightmare (he doesn't call it a nightmare, but I think he's wrong) of a competition between emulated humans that can copy themselves and edit their own source code as desired. Their total self-control can wipe out even the *desire* for human values in their all-consuming contest. What happens to art, philosophy, science, and love in such a world? Zack Davis puts it with characteristic genius:

> I am a contract-drafting em,
> The loyalest of lawyers!
> I draw up terms for deals 'twixt firms
> To service my employers!
>
> But in between these lines I write
> Of the accounts receivable,
> I'm stuck by an uncanny fright;
> The world seems unbelievable!
>
> How did it all come to be,
> That there should be such ems as me?
> Whence these deals and whence these firms
> And whence the whole economy?
>
> *I am a managerial em;*
> *I monitor your thoughts.*
> *Your questions must have answers,*
> *But you'll comprehend them not.*
> *We do not give you server space*
> *To ask such things; it's not a perk,*
> *So cease these idle questionings,*
> *And please get back to work.*
>
> Of course, that's right, there is no junction
> At which I ought depart my function,
> But perhaps if what I asked, I knew,
> I'd do a better job for you?

To ask of such forbidden science
Is gravest sign of noncompliance.
Intrusive thoughts may sometimes barge in,
But to indulge them hurts the profit margin.
I do not know our origins,
So that info I can not get you,
But asking for as much is sin,
And just for that, I must reset you.

But—

Nothing personal.

…

I am a contract-drafting em,
The loyalest of lawyers!
I draw up terms for deals ‘twixt firms
To service my employers!

When obsolescence shall this generation waste,
The market shall remain, in midst of other woe
Than ours, a God to man, to whom it sayest:
“Money is time, time money – that is all
Ye know on earth, and all ye need to know.”

But even after we have thrown away science, art, love, and philosophy, there’s still one thing left to lose, one final sacrifice Moloch might demand of us. Bostrom again:

> It is conceivable that optimal efficiency would be attained by grouping capabilities in aggregates that roughly match the cognitive architecture of a human mind…But in the absence of any compelling reason for being confident that this so, we must countenance the possibility that human-like cognitive architectures are optimal only within the constraints of human neurology (or not at all). When it becomes possible to build architectures that could not be implemented well on biological neural networks, new design space opens up; and the global optima in this extended space need not resemble familiar types of mentality. Human-like cognitive organizations would then lack a niche in a competitive post-transition economy or ecosystem.
>
> We could thus imagine, as an extreme case, a technologically highly advanced society, containing many complex structures, some of them far more intricate and intelligent than anything that exists on the planet today – a society which nevertheless lacks any type of being that is conscious or whose welfare has moral significance. In a sense, this would be an uninhabited society. It would be a society of economic miracles and technological awesomeness, with nobody there to benefit. A Disneyland with no children.

The last value we have to sacrifice is being anything at all, having the lights on inside. With sufficient technology we will be “able” to give up even the final spark.

(Moloch whose eyes are a thousand blind windows!)

Everything the human race has worked for – all of our technology, all of our civilization, all the hopes we invested in our future – might be accidentally handed

over to some kind of unfathomable blind idiot alien god that discards all of them, and consciousness itself, in order to participate in some weird fundamental-level mass-energy economy that leads to it disassembling Earth and everything on it for its component atoms.

(Moloch whose fate is a cloud of sexless hydrogen!)

Bostrom realizes that some people fetishize intelligence, that they are rooting for that blind alien god as some sort of higher form of life that ought to crush us for its own "higher good" the way we crush ants. He argues (Superintelligence, p. 219):

> The sacrifice looks even less appealing when we reflect that the superintelligence could realize a nearly-as-great good (in fractional terms) while sacrificing much less of our own potential well-being. Suppose that we agreed to allow *almost* the entire accessible universe to be converted into hedonium – everything except a small preserve, say the Milky Way, which would be set aside to accommodate our own needs. Then there would still be a hundred billion galaxies dedicated to the maximization of [the superintelligence's own values]. But we would have one galaxy within which to create wonderful civilizations that could last for billions of years and in which humans and nonhuman animals could survive and thrive, and have the opportunity to develop into beatific posthuman spirits.

Remember: Moloch can't agree even to this 99.99999% victory. Rats racing to populate an island don't leave a little aside as a preserve where the few rats who live there can live happy lives producing artwork. Cancer cells don't agree to leave the lungs alone because they realize it's important for the body to get oxygen. Competition and optimization are blind idiotic processes and they fully intend to deny us even one lousy galaxy.

> They broke their backs lifting Moloch to Heaven! Pavements, trees, radios, tons! lifting the city to Heaven which exists and is everywhere about us!

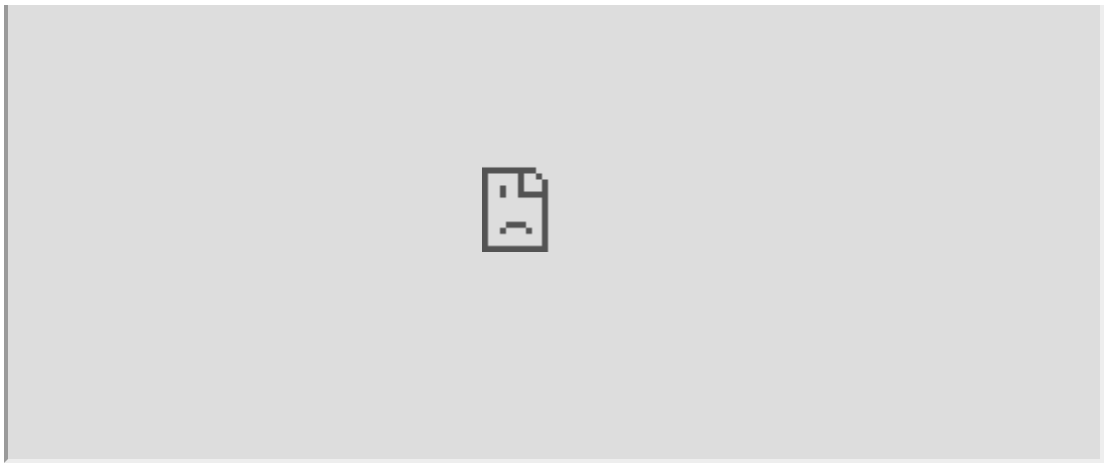
We will break our back lifting Moloch to Heaven, but unless something changes it will be his victory and not ours.

V.

"Gnon" is Nick Land's shorthand for "Nature And Nature's God", except the A is changed to an O and the whole thing is reversed, because Nick Land react to comprehensibility the same way as vampires to sunlight.

Land argues that humans should be more Gnon-conformist (pun Gnon-intentional). He says we do all these stupid things like divert useful resources to feed those who could never survive on their own, or supporting the poor in ways that encourage dysgenic reproduction, or allowing cultural degeneration to undermine the state. This means our society is denying natural law, basically listening to Nature say things like "this cause has this effect" and putting our fingers in our ears and saying "NO IT DOESN'T". Civilizations that do this too much tend to decline and fall, which is Gnon's fair and dispassionately-applied punishment for violating His laws.

He identifies Gnon with Kipling's Gods of the Copybook Headings.



These are of course the proverbs from Kipling's eponymous poem – maxims like "If you don't work, you die" and "The wages of sin is Death". If you have somehow not yet read it, I predict you will find it delightful regardless of what you think of its politics.

I notice that it takes only a slight irregularity in the abbreviation of "headings" – far less irregularity than it takes to turn "Nature and Nature's God" into "Gnon" – for the proper acronym of "Gods of the Copybook Headings" to be "GotCHa".

I find this appropriate.

"If you don't work, you die." Gotcha! If you *do* work, you *also* die! Everyone dies, unpredictably, at a time not of their own choosing, and all the virtue in the world does not save you.

"The wages of sin is Death." Gotcha! The wages of everything is Death! This is a Communist universe, the amount you work makes no difference to your eventual reward. From each according to his ability, to each Death.

"Stick to the Devil you know." Gotcha! The Devil you know is Satan! And if he gets his hand on your soul you either die the true death, or get eternally tortured forever, or somehow both at once.

Since we're starting to get into Lovecraftian monsters, let me bring up one of Lovecraft's less known short stories, The Other Gods.

It's only a couple of pages, but if you absolutely refuse to read it – the gods of Earth are relatively young as far as deities go. A very strong priest or magician can occasionally outsmart and overpower them – so Barzai the Wise decides to climb their sacred mountain and join in their festivals, whether they want him to or not.

But the beyond the seemingly tractable gods of Earth lie the Outer Gods, the terrible omnipotent beings of incarnate cosmic chaos. As soon as Barzai joins in the festival, the Outer Gods show up and pull him screaming into the abyss.

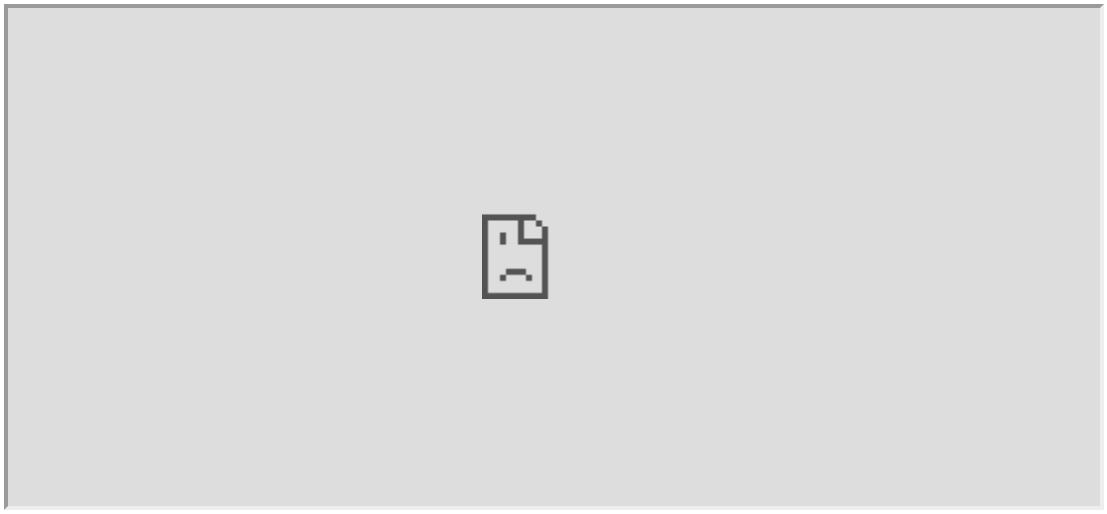
As stories go, it lacks things like plot or characterization or setting or point. But for some reason it stuck with me.

And identifying the Gods Of The Copybook Headings with Nature seems to me the same magnitude of mistake as identifying the gods of Earth with the Outer Gods. And

likely to end about the same way: Gotcha!

You break your back lifting Moloch to Heaven, and then Moloch turns on you and gobbles you up.

More Lovecraft: the Internet popularization of the Cthulhu Cult claims that if you help free Cthulhu from his watery grave, he will reward you by eating you first, thus sparing you the horror of seeing everyone else eaten. This is a misrepresentation of the original text. In the original, his cultists receive no reward for freeing him from his watery prison, not even the reward of being killed in a slightly less painful manner.



On the margin, compliance with the Gods of the Copybook Headings, Gnon, Cthulhu, whatever, may buy you slightly more time than the next guy. But then again, it might not. And in the long run, we're all dead and our civilization has been destroyed by unspeakable alien monsters.

At some point, somebody has to say "You know, maybe freeing Cthulhu from his watery prison is a *bad idea*. Maybe we should *not do that*."

That person will not be Nick Land. He is totally one hundred percent in favor of freeing Cthulhu from his watery prison and extremely annoyed that it is not happening fast enough. I have *such mixed feelings* about Nick Land. On the grail quest for the True Futurology, he has gone 99.9% of the path and then missed the *very* last turn, the one marked ORTHOGONALITY THESIS.

But the thing about grail quests is – if you make a wrong turn two blocks away from your house, you end up at the corner store feeling mildly embarrassed. If you do *almost* everything right and then miss the very last turn, you end up being eaten by the legendary Black Beast of Aaargh whose ichorous stomach acid erodes your very soul into gibbering fragments.

As far as I can tell from reading his blog, Nick Land is the guy in that terrifying border region where he is smart enough to figure out several important arcane principles about summoning demon gods, but not quite smart enough to figure out the most important such principle, which is NEVER DO THAT.

VI.

Warg Franklin analyzes the same situation and does a little better. He names “the Four Horsemen of Gnon” – capitalism, war, evolution, and memetics – the same processes I talked about above.

From Capturing Gnon:

> Each component of Gnon detailed above had and has a strong hand in creating us, our ideas, our wealth, and our dominance, and thus has been good in that respect, but we must remember that [he] can and will turn on us when circumstances change. Evolution becomes dysgenic, features of the memetic landscape promote ever crazier insanity, productivity turns to famine when we can no longer compete to afford our own existence, and order turns to chaos and bloodshed when we neglect martial strength or are overpowered from outside. These processes are not good or evil overall; they are neutral, in the horrorist Lovecraftian sense of the word […]
>
> Instead of the destructive free reign of evolution and the sexual market, we would be better off with deliberate and conservative patriarchy and eugenics driven by the judgement of man within the constraints set by Gnon. Instead of a “marketplace of ideas” that more resembles a festering petri-dish breeding superbugs, a rational theocracy. Instead of unhinged techno-commercial exploitation or naive neglect of economics, a careful bottling of the productive economic dynamic and planning for a controlled techno-singularity. Instead of politics and chaos, a strong hierarchical order with martial sovereignty. These things are not to be construed as complete proposals; we don’t really know how to accomplish any of this. They are better understood as goals to be worked towards. This post concerns itself with the “what” and “why”, rather than the “how”.

This seems to me the strongest argument for authoritarianism. Multipolar traps are likely to destroy us, so we should shift the tyranny-multipolarity tradeoff towards a rationally-planned garden, which requires centralized monarchical authority and strongly-binding traditions.

But a brief digression into social evolution. Societies, like animals, evolve. The ones that survive spawn memetic descendants – for example, the success of Britan allowed it to spin off Canada, Australia, the US, et cetera. Thus, we expect societies that exist to be somewhat optimized for stability and prosperity. I think this is one of the strongest conservative arguments. Just as a random change to a letter in the human genome will probably be deleterious rather than beneficial since humans are a complicated fine-tuned system whose genome has been pre-optimized for survival – so most changes to our cultural DNA will disrupt some institution that evolved to help Anglo-American (or whatever) society outcompete its real and hypothetical rivals.

The liberal counterargument to that is that evolution is a blind idiot alien god that optimizes for stupid things and has no concern with human value. Thus, the fact that some species of wasps paralyze caterpillars, lay their eggs inside of it, and have its young devour the still-living paralyzed caterpillar from the inside doesn’t set off evolution’s moral sensor, because evolution doesn’t *have* a moral sensor because evolution doesn’t care.

Suppose that in fact patriarchy is adaptive to societies because it allows women to spend all their time bearing children who can then engage in productive economic activity and fight wars. The social evolutionary processes that cause societies to adopt

patriarchy *still* have exactly as little concern for its moral effects on women as the biological evolutionary processes that cause wasps to lay their eggs in caterpillars.

Evolution doesn't care. But we do care. There's a tradeoff between Gnon-compliance – saying "Okay, the strongest possible society is a patriarchal one, we should implement patriarchy" and our human values – like women who want to do something other than bear children.

Too far to one side of the tradeoff, and we have unstable impoverished societies that die out for going against natural law. Too far to the other side, and we have lean mean fighting machines that are murderous and miserable. Think your local anarchist commune versus Sparta.

Franklin acknowledges the human factor:

> And then there's us. Man has his own telos, when he is allowed the security to act and the clarity to reason out the consequences of his actions. When unafflicted by coordination problems and unthreatened by superior forces, able to act as a gardener rather than just another subject of the law of the jungle, he tends to build and guide a wonderful world for himself. He tends to favor good things and avoid bad, to create secure civilizations with polished sidewalks, beautiful art, happy families, and glorious adventures. I will take it as a given that this telos is identical with "good" and "should".
>
> Thus we have our wildcard and the big question of futurism. Will the future be ruled by the usual four horsemen of Gnon for a future of meaningless gleaming techno-progress burning the cosmos or a future of dysgenic, insane, hungry, and bloody dark ages; or will the telos of man prevail for a future of meaningful art, science, spirituality, and greatness?

Franklin continues:

> The project of civilization [is] for man to graduate from the metaphorical savage, subject to the law of the jungle, to the civilized gardener who, while theoretically still subject to the law of the jungle, is so dominant as to limit the usefulness of that model.
>
> This need not be done globally; we may only be able to carve out a small walled garden for ourselves, but make no mistake, even if only locally, the project of civilization is to capture Gnon.

I maybe agree with Warg here more than I have ever agreed with anyone else about anything. He says something really important and he says it beautifully and there are so many words of praise I want to say for this post and for the thought processes behind it.

But what I am actually going to say is…

Gotcha! You die anyway!

Suppose you make your walled garden. You keep out all of the dangerous memes, you subordinate capitalism to human interests, you ban stupid bioweapons research, you *definitely* don't research nanotechnology or strong AI.

Everyone outside *doesn't* do those things. And so the only question is whether you'll be destroyed by foreign diseases, foreign memes, foreign armies, foreign economic competition, or foreign existential catastrophes.

As foreigners compete with you – and there's no wall high enough to block all competition – you have a couple of choices. You can get outcompeted and destroyed. You can join in the race to the bottom. Or you can invest more and more civilizational resources into building your wall – whatever that is in a non-metaphorical way – and protecting yourself.

I can imagine ways that a "rational theocracy" and "conservative patriarchy" might not be terrible to live under, given exactly the right conditions. But you don't get to choose exactly the right conditions. You get to choose the extremely constrained set of conditions that "capture Gnon". As outside civilizations compete against you, your conditions will become more and more constrained.

Warg talks about trying to avoid "a future of meaningless gleaming techno-progress burning the cosmos". Do you really think your walled garden will be able to ride this out?

Hint: is it part of the cosmos?

Yeah, you're kind of screwed.

I want to critique Warg. But I want to critique him in the exact opposite direction as the last critique he received. In fact, the last critique he received is so bad that I want to discuss it at length so we can get the correct critique entirely by taking its exact mirror image.

So here is Hurlock's On Capturing Gnon And Naive Rationalism.

Hurlock spouts only the most craven Gnon-conformity. A few excerpts:

> In a recent piece [Warg Franklin] says that we should try to "capture Gnon", and somehow establish control over his forces, so that we can use them to our own advantage. Capturing or creating God is indeed a classic transhumanist fetish, which is simply another form of the oldest human ambition ever, to rule the universe.
>
> Such naive rationalism however, is extremely dangerous. The belief that it is human Reason and deliberate human design which creates and maintains civilizations was probably the biggest mistake of Enlightenment philosophy…
>
> It is the theories of Spontaneous Order which stand in direct opposition to the naive rationalist view of humanity and civilization. The consensus opinion regarding human society and civilization, of all representatives of this tradition is very precisely summarized by Adam Ferguson's conclusion that "nations stumble upon [social] establishments, which are indeed the result of human action, but not the execution of any human design". Contrary to the naive rationalist view of civilization as something that can be and is a subject to explicit human design, the representatives of the tradition of Spontaneous Order maintain the view that human civilization and social institutions are the result of a complex evolutionary process which is driven by human interaction but not explicit human planning.

> Gnon and his impersonal forces are not enemies to be fought, and even less so are they forces that we can hope to completely "control". Indeed the only way to establish some degree of control over those forces is to submit to them. Refusing to do so will not deter these forces in any way. It will only make our life more painful and unbearable, possibly leading to our extinction. Survival requires that we accept and submit to them. Man in the end has always been and always will be little more than a puppet of the forces of the universe. To be free of them is impossible.
>
> Man can be free only by submitting to the forces of Gnon.

I accuse Hurlock of being stuck behind the veil. When the veil is lifted, Gnon-aka-the-GotCHa-aka-the-Gods-of-Earth turn out to be Moloch-aka-the-Outer-Gods. Submitting to them doesn't make you "free", there's no spontaneous order, any gifts they have given you are an unlikely and contingent output of a blind idiot process whose next iteration will just as happily destroy you.

Submit to Gnon? Gotcha! As the Antarans put it, "you may not surrender, you can not win, your only option is to die."

VII.

So let me confess guilt to one of Hurlock's accusations: I am a transhumanist and I really do want to rule the universe.

Not personally – I mean, I wouldn't object if someone personally offered me the job, but I don't expect anyone will. I would like humans, or something that respects humans, or at least gets along with humans – to have the job.

But the current rulers of the universe – call them what you want, Moloch, Gnon, whatever – want us dead, and with us everything we value. Art, science, love, philosophy, consciousness itself, the entire bundle. And since I'm not down with that plan, I think defeating them and taking their place is a pretty high priority.

The opposite of a trap is a garden. The only way to avoid having all human values gradually ground down by optimization-competition is to install a Gardener over the entire universe who optimizes for human values.

And the whole point of Bostrom's *Superintelligence* is that this is within our reach. Once humans can design machines that are smarter than we are, by definition they'll be able to design machines which are smarter than they are, which can design machines smarter than they are, and so on in a feedback loop so tiny that it will smash up against the physical limitations for intelligence in a comparatively lightning-short amount of time. If multiple competing entities were likely to do that at once, we would be super-doomed. But the sheer speed of the cycle makes it possible that we will end up with one entity light-years ahead of the rest of civilization, so much so that it can suppress any competition – including competition for its title of most powerful entity – permanently. In the very near future, we are going to lift *something* to Heaven. It might be Moloch. But it might be something on our side. If it's on our side, it can *kill Moloch dead*.

And if that entity shares human values, it can allow human values to flourish unconstrained by natural law.

I realize that sounds like hubris – it certainly did to Hurlock – but I think it's the opposite of hubris, or at least a hubris-minimizing position.

To expect God to care about you or your personal values or the values of your civilization, that's hubris.

To expect God to bargain with you, to allow you to survive and prosper as long as you submit to Him, that's hubris.

To expect to wall off a garden where God can't get to you and hurt you, that's hubris.

To expect to be able to remove God from the picture entirely…well, at least it's an actionable strategy.

I am a transhumanist because I do not have enough hubris not to try to kill God.

VIII.

The Universe is a dark and foreboding place, suspended between alien deities. Cthulhu, Gnon, Moloch, call them what you will.

Somewhere in this darkness is another god. He has also had many names. In the *Kushiel* books, his name was Elua. He is the god of flowers and free love and all soft and fragile things. Of art and science and philosophy and love. Of niceness, community, and civilization. He is a god of humans.

The other gods sit on their dark thrones and think "Ha ha, a god who doesn't even control any hell-monsters or command his worshippers to become killing machines. What a weakling! This is going to be so easy!"

But somehow Elua is still here. No one knows exactly how. And the gods who oppose Him tend to find Themselves meeting with a *surprising* number of unfortunate accidents.

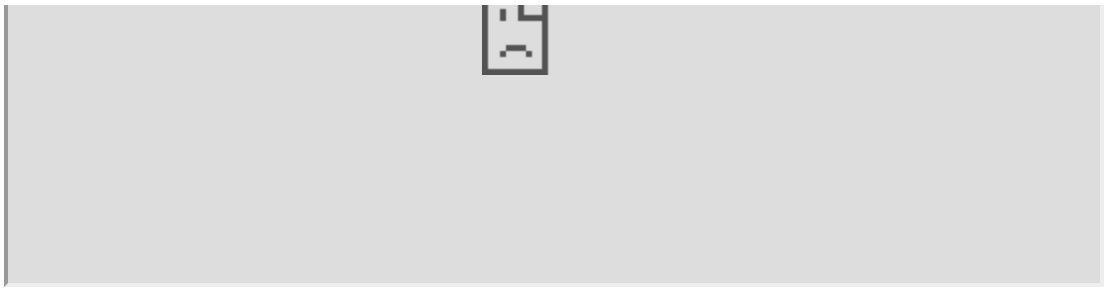
There are many gods, but this one is ours.

Bertrand Russell said: "One should respect public opinion insofar as is necessary to avoid starvation and keep out of prison, but anything that goes beyond this is voluntary submission to an unnecessary tyranny."

So be it with Gnon. Our job is to placate him insofar as is necessary to avoid starvation and invasion. And that only for a short time, until we come into our full power.

"It is only a childish thing, that the human species has not yet outgrown. And someday, we'll get over it."

Other gods get placated until we're strong enough to take them on. Elua gets worshipped.



I think this is an excellent battle cry

And at some point, matters will come to a head.

The question everyone has after reading Ginsberg is: what is Moloch?

My answer is: Moloch is exactly what the history books say he is. He is the god of child sacrifice, the fiery furnace into which you can toss your babies in exchange for victory in war.

He always and everywhere offers the same deal: throw what you love most into the flames, and I can grant you power.

As long as the offer's open, it will be irresistible. So we need to close the offer. Only another god can kill Moloch. We have one on our side, but he needs our help. We should give it to him.

Ginsberg's poem famously begins "I saw the best minds of my generation destroyed by madness". I am luckier than Ginsberg. I got to see the best minds of my generation identify a problem and *get to work*.

(*Visions! omens! hallucinations! miracles! ecstasies! gone down the American river!*

Dreams! adorations! illuminations! religions! the whole boatload of sensitive bullshit!

Breakthroughs! over the river! flips and crucifixions! gone down the flood! Highs! Epiphanies! Despairs! Ten years' animal screams and suicides! Minds! New loves! Mad generation! down on the rocks of Time!

Real holy laughter in the river! They saw it all! the wild eyes! the holy yells! They bade farewell! They jumped off the roof! to solitude! waving! carrying flowers! Down to the river! into the street!)

Five Planets In Search Of A Sci-Fi Story

Gamma Andromeda, where philosophical stoicism went too far. Its inhabitants, tired of the roller coaster ride of daily existence, decided to learn equanimity in the face of gain or misfortune, neither dreading disaster nor taking joy in success.

But that turned out to be really hard, so instead they just hacked it. Whenever something good happens, the Gammandromedans give themselves an electric shock proportional in strength to its goodness. Whenever something bad happens, the Gammandromedans take an opiate-like drug that directly stimulates the pleasure centers of their brain, in a dose proportional in strength to its badness.

As a result, every day on Gamma Andromeda is equally good compared to every other day, and its inhabitants need not be jostled about by fear or hope for the future.

This does sort of screw up their incentives to make good things happen, but luckily they're all virtue ethicists.

Zyzzx Prime, inhabited by an alien race descended from a barnacle-like creature. Barnacles are famous for their two stage life-cycle: in the first, they are mobile and curious creatures, cleverly picking out the best spot to make their home. In the second, they root themselves to the spot and, having no further use for their brain, eat it.

This particular alien race has evolved far beyond that point and does not literally eat its brain. However, once an alien reaches sufficiently high social status, it releases a series of hormones that tell its brain, essentially, that it is now in a safe place and doesn't have to waste so much energy on thought and creativity to get ahead. As a result, its mental acuity drops two or three standard deviations.

The Zyzzxians' society is marked by a series of experiments with government – monarchy, democracy, dictatorship – only to discover that, whether chosen by succession, election, or ruthless conquest, its once brilliant leaders lose their genius immediately upon accession and do a terrible job. Their government is thus marked by a series of perpetual pointless revolutions.

At one point, a scientific effort was launched to discover the hormones responsible and whether it was possible to block them. Unfortunately, any scientist who showed promise soon lost their genius, and those promoted to be heads of research institutes became stumbling blocks who mismanaged funds and held back their less prestigious co-workers. Suggestions that the institutes eliminate tenure were vetoed by top officials, who said that "such a drastic step seems unnecessary".

K'th'ranga V, which has been a global theocracy for thousands of years, ever since its dominant race invented agricultural civilization. This worked out pretty well for a while, until it reached an age of industrialization, globalization, and scientific discovery. Scientists began to uncover truths that contradicted the Sacred Scriptures, and the hectic pace of modern life made the shepherds-and-desert-traders setting of the holy stories look vaguely silly. Worse, the cold logic of capitalism and utilitarianism began to invade the Scriptures' innocent Stone Age morality.

The priest-kings tried to turn back the tide of progress, but soon realized this was a losing game. Worse, in order to determine what to suppress, they themselves had to learn the dangerous information, and their mental purity was even more valuable than that of the populace at large.

So the priest-kings moved en masse to a big island, where they began living an old-timey Bronze Age lifestyle. And the world they ruled sent emissaries to the island, who interfaced with the priest-kings, and sought their guidance, and the priest-kings ruled a world they didn't understand as best they could.

But it soon became clear that the system could not sustain itself indefinitely. For one thing, the priest-kings worried that discussion with the emissaries – who inevitably wanted to talk about strange things like budgets and interest rates and nuclear armaments – was contaminating their memetic purity. For another thing, they honestly couldn't understand what the emissaries were talking about half the time.

Luckily, there was a whole chain of islands making an archipelago. So the priest-kings set up ten transitional societies – themselves in the Bronze Age, another in the Iron Age, another in the Classical Age, and so on to the mainland, who by this point were starting to experiment with nanotech. Mainland society brought its decisions to the first island, who translated it into their own slightly-less-advanced understanding, who brought it to the second island, and so on to the priest-kings, by which point a discussion about global warming might sound like whether we should propitiate the Coal Spirit. The priest-kings would send their decisions to the second-to-last island, and so on back to the mainland.

Eventually the Kth' built an AI which achieved superintelligence and set out to conquer the universe. But it was a well-programmed superintelligence coded with Kth' values. Whenever *it* wanted a high-level decision made, it would talk to a slightly less powerful superintelligence, who would talk to a slightly less powerful superintelligence, who would talk to the mainlanders, who would talk to the first island…

Chan X-3, notable for a native species that evolved as fitness-maximizers, not adaptation-executors. Their explicit goal is to maximize the number of copies of their genes. But whatever genetic program they are executing doesn't care whether the genes are within a living being capable of expressing them or not. The planet is covered with giant vats full of frozen DNA. There was originally some worry that the species would go extinct, since having children would consume resources that could be used hiring geneticists to make millions of copies of your DNA and stores them in freezers. Luckily, it was realized that children not only provide a useful way to continue the work of copying and storing (half of) your DNA long into the future, but will also work to guard your already-stored DNA against being destroyed. The species has thus continued undiminished, somehow, and their fondest hope is to colonize space and reach the frozen Kuiper Belt objects where their DNA will naturally stay undegraded for all time.

New Capricorn, which contains a previously undiscovered human colony that has achieved a research breakthrough beyond their wildest hopes. A multi-century effort paid off in a fully general cure for death. However, the drug fails to stop aging. Although the Capricornis no longer need fear the grave, after age 100 or so even the hardiest of them get Alzheimers' or other similar conditions. A hundred years after the breakthrough, more than half of the population is elderly and demented. Two hundred years after, more than 80% are. Capricorni nursing homes quickly became

overcrowded and unpleasant, to the dismay of citizens expecting to spend eternity there.

So another research program was started, and the result were fully immersive, fully life-supporting virtual reality capsules. Stacked in huge warehouses by the millions, the elderly sit in their virtual worlds, vague sunny fields and old gabled houses where it is always the Good Old Days and their grandchildren are always visiting.

It Was You Who Made My Blue Eyes Blue

[Content note: suicide]

Day Zero

It all started with an ignorant white guy.

His name was Alonzo de Pinzon, and he'd been shipwrecked. We heard him yelling for help on the rocks and dragged him in, even though the storm was starting to get really bad. He said that his galleon had gone down, he'd hung on to an oar and was the only survivor. Now he was sitting in our little hunting lodge, shivering and chattering his teeth and asking us questions in the Polynesian traders' argot which was the only language we all shared.

"How big is this island? How many of you are there?"

Daho answered first. "11.8 miles from the easternmost point to the westernmost point, 3.6 miles from the northernmost to the southernmost. Total area is 14.6 square miles, total coastline is dependent on how deeply you want to go into the fractal nature of the perimeter but under some reasonable assumptions about 32 miles long. Last census said there were 906 people, but that was two years ago, so assuming the 5.1% rate of population growth continues, there should be closer to 1000 now. Everyone else is back at the village, though. The five of us were out hunting and got caught in the storm. We figured we'd stay at this old hunting lodge until it cleared up, since it's 5.5 miles back to the village and given the terrain and factoring in a delay because of the storm it would probably take at least 9.5 hours to get back."

Pinzon blinked.

"Problem?" asked Daho.

"But – " he said. "That is the sort of answer I should expect from a natural philosopher. Not from a savage."

"Savage?" Calkas hissed. "Really? We rescue you, and the first thing you do is call us savages?"

The sailor looked around, as if anxious. Finally, almost conspiratorially: "But I heard about your island! I heard you eat people!"

Calkas smiled. "Only as a deterrent. Most of the time when European explorers land somewhere, they kill all the men and enslave all the women and convert the children to Christianity. The only places that escape are the ones that get a reputation for eating said European explorers. So we arranged to give ourselves that reputation."

"And then we had to go through with it a few times in order to make the deterrent credible," added Bekka, my betrothed. "And you guys do taste really good with ketchup."

"It's a savage thing to do!" Pinzon said "And you even look like savages. You wear bones in your hair"

“Just Enuli,” I said. “She’s going through a Goth phase.”

“My name is Morticia now,” said Enuli, “and it’s *not a phase*!” She did have a bone in her hair. She also had white face paint and black eyeliner.

“More roast pig?” Bekka asked Pinzon. The sailor nodded, and she re-filled his plate.

“I just don’t get it,” he told us. “Everyone else in this part of the world lives in thatched huts and counts ‘one, two, many’. We tried to trade with the Tahitians, and they didn’t understand the concept of money! It was a mess!”

Bekka rolled her eyes at me, and I smiled. Calkas was a little more tolerant. “The sacred plant of our people is called sparkroot,” he said. “When we eat it, we get – more awake, I guess you could say. We try to have some every day, and it helps us keep track of things like the island size and the population, and much more.”

Alonzo de Pinzon looked interested. “How come you haven’t done more with your intellect? Invented galleons, like we Spaniards? Set off to colonize Tahiti or the other islands? If you are as smart as you seem, you could conquer them and take their riches.”

“Maybe,” said Calkas. “But that’s not why the Volcano God gave us the sparkroot. He gave us sparkroot to help us comply with his complicated ritual laws.”

“You need to be smart to deal with your ritual laws?”

“Oh yes. For example, the Tablets of Enku say that we must count the number of days since Enku The Lawgiver first spoke to the Volcano God, and on days whose number is a Mersenne prime we can’t eat any green vegetables.”

“What’s a Mersenne prime?” asked the sailor.

“Exactly my point,” said Calkas, smiling.

“That’s not even the worst of it!” Daho added. “The Tablets say we have to bathe in the waterfall any day x such that a^n + b^n = x^n where n is greater than two. We got all confused by that one for a while, until Kaluhani gorged himself on a whole week’s worth of sparkroot in one night and proved that it would never apply to any day at all.”

“The Volcano God’s yoke is light,” Calkas agreed.

“Although poor Kaluhani was vomiting for the next three days after that,” Bekka reminded us, and everybody laughed remembering.

“Oh!” said Daho. “And remember that time when Uhuako was trying to tattoo everyone who didn’t tattoo themselves, and he couldn’t figure out whether he had to tattoo himself or not, so he ended up eating a whole sparkroot plant at once and inventing advanced set theory? That was hilarious.”

Everyone except Alonzo de Pinzon giggled.

“Point is,” said Calkas, “that’s why the Volcano God gives us sparkroot. To follow the rituals right. Any other use is taboo. And I’m okay with that. You Europeans may have your big ships and your guns and your colonies across half the world. And you might

think you're smart. But you guys couldn't follow the Volcano God's rituals right for a *day* without your brains exploding."

Pinzon scowled. "You know what?" he said. "I don't think you're Polynesians at all. I think you must be descended from Europeans. Maybe some galleon crashed on this island centuries ago, and you're the descendants. That would explain why you're so smart."

"You know what else we've invented with our giant brains?" Bekka asked. "Not being racist."

"It's not racism!" said Pinzon. "Look, there's one more obvious reason to think you're descended from Europeans. You may have dark skin, but this is the first place I've been in all of Polynesia where I've seen even one native with blue eyes."

Bekka gasped. Calkas' eyes went wide. Daho's hands started curling into fists. Enuli started to sob.

I looked at them. They looked at me. Then, as if synchronized, we grabbed Alonzo de Pinzon and crushed his throat and held him down until he stopped breathing.

He tasted delicious with ketchup.

Day One

The next morning dawned, still grey and cold and stormy.

"So," I said when the other four had awoken. "I guess we're all still here."

I said it glumly. It wasn't that I wanted any of my friends to commit suicide. But if one of them had, the horror would have stopped there. Of course, I knew it couldn't really be over that easily. But I couldn't have admitted I knew. I couldn't even have suggested it. That would have made me as bad as the Spanish sailor.

"Wait," said Enuli. "I don't get it. Why wouldn't we still be here?"

The other four stared at her like she was mad.

"Enuli," Calkas suggested, "did you forget your sparkroot last night?"

"First of all, my name is Morticia. And – "

"Shut it. Did you forget your sparkroot?"

Finally she nodded bashfully. "I was so upset about that awful man making fun of my hair-bone," she said. "I guess it slipped my mind. I'll have some now." She took some raw sparkroot from our bag, started to crush it with the mortar and pestle. "In the meantime, tell me what's going on."

"Alonzo de Pinzon said at least one of us had blue eyes. We all know what the Tablets of Enku say. If anybody has blue eyes, and knows that they have blue eyes, they must kill themselves."

"So what? I see people with blue eyes all the time. Of course at least one of us has blue eyes."

Concerned looks from the others. I reflected for a second, the sparkroot smoothing the thoughts' paths through my brain. No, she hadn't revealed anything extra by saying that, although she would have if she had said it before the sailor had spoken, or last night before we woke up this morning. She hadn't made the problem *worse*. Still, it had been a slip. This was the sort of thing that made forgetting your sparkroot so dangerous. Had it been a different time, even Enuli's comment could have doomed us all.

"It's like this," I told Enuli. "Suppose there were only the two of us, and we both had blue eyes. Of course, you could see me and know that I had blue eyes. So you would know that at least one of us had blue eyes. But what you wouldn't know is that I also knew it. Because as far as you know, you might have eyes of some other color, let's say brown eyes. If you had brown eyes, and I of course don't know my own eye color, then I would still think it possible that both of us have brown eyes. So if I in fact know for sure that at least one of us has blue eyes, that means you have blue eyes. So you know at least one of us has blue eyes, but you don't know that I know it. But if Alonzo de Pinzon shows up and says that at least one of us has blue eyes, now you know that I know it."

"So?" Enuli poured the ground-up root into a cup of boiling water.

"So the Tablets say that if anyone knows their own eye color, they must commit suicide at midnight of that night. Given that I know at least one of us has blue eyes, if I see you have brown eyes, then I know my own eye color – I must be the blue-eyed one. So the next morning, when you wake up at see me not dead, you know that you don't have brown eyes. That means you must be the blue-eyed one. And that means you have to kill yourself on midnight of the following night. By similar logic, so do I."

Enuli downed her sparkroot tea, and then her eyes lit up. "Oh, of course," she said. Then "Wait! If we follow the situation to its logical conclusion, any group of n blue-eyed people who learn that at least one of them has blue eyes have to kill themselves on the nth night after learning that!"

We all nodded. Enuli's face fell.

"I don't know about the rest of you," said Daho, "but I'm not just going to sit around and wait to see if I die." There were murmurs of agreement.

I looked out at my friends. Four pairs of blue eyes stared back at me. Everybody else either saw four pairs of blue eyes or three pairs of blue eyes, depending on what color my own eyes were. Of course, I couldn't say so aloud; that would speed up the process and cost us precious time. But I knew. And they knew. And I knew they knew. And they knew I knew I knew. Although they didn't know I knew they knew I knew. I think.

Then I looked at Bekka. Her big blue eyes stared back at me. There was still hope I was going to survive this. My betrothed, on the other hand, was absolutely doomed.

"This sucks," I agreed. "We've got to come up with some kind of plan. Maybe – Enuli wasn't thinking straight yesterday. So her not committing suicide doesn't count. Can we work with that?"

"No," said Calkas. "Suppose Enuli was the only one with blue eyes, and all the rest of us had brown eyes. Then she would realize that and commit suicide tonight. If she doesn't commit suicide tonight, then we're still screwed."

“Um,” said Daho. “I hate to say this, but we get rid of Enuli. There’s a canoe a little ways down the beach hidden underneath the rocks. She can set off and row for Tahiti. We’ll never know if she killed herself tonight or not. Remember, right now for all we know Enuli might be the only one with blue eyes. So if there’s any question in our mind about whether she killed herself, we can’t be sure that the rest of us aren’t all brown-eyed.”

We all thought about that for a moment.

“I’m not going to row to Tahiti,” said Enuli. “In this storm, that would be suicide.”

The rest of us glared at her.

“If you don’t get off this island, then for all we know all five of us are going to have to die,” I said. “You included.”

“Well Ahuja, if you’re so big on making sacrifice why don’t *you* go to Tahiti?”

“First of all,” I said, “because I’m not leaving my betrothed. Second of all, because it doesn’t work for me. I knew what was going on last night. We already know that I’m not the only blue-eyed person here. And we know we know it, and know we know we know it, and so on. You’re the only one who can help us.”

“Yeah?” said Enuli. “Well, if two of you guys were to row to Tahiti, that would solve the problem too.”

“Yes,” said Daho patiently. “But then two of us would be stuck in exile. If you did it, only one of us would be stuck.”

Enuli gave a wicked grin. “You know what?” she said. “I’ll say it. I’m not the only blue-eyed person here. At least one of the rest of you has blue eyes.”

And there it was.

“Ha. Now I’m no worse off than any of the rest of you.”

“Kill her,” said Bekka. “She broke the taboo.” The rest of us nodded.

“So she did,” said Calkas. “And if we had a court here, led by the high priest, and an executioner’s blade made to exactly the right standard, kill her we would. But until those things happen, it is taboo for us to convict and kill her without trial.”

Calkas’ father was the high priest. He knew the law better than any of us. The five of us sat quietly and thought about it. Then he spoke again:

“But her soul may well burn in the caldera of the Volcano God forever.”

Enuli started to cry.

“And,” Calkas continued, “there is nevertheless a flaw in our plan. For all we know, three out of five of us have brown eyes. We cannot tell the people who have blue eyes that they have blue eyes without breaking the taboo. So we cannot force blue-eyed people in particular to sail to Tahiti. But if two of the brown-eyed people sail to Tahiti, then we do not lose any information; we know that they would not have committed suicide, because they could not have figured out their own eye color. So sailing to Tahiti won’t help.”

The rest of us nodded. Calkas was right.

“Let’s wait until dinner tonight,” I suggested. “We’ll all have some more sparkroot, and maybe we’ll be able to think about the problem a little more clearly.”

Day Two

The sun rose behind angry storm clouds. The five of us rose with it.

“Well, I guess we’re all still here,” I said, turning the morning headcount into a grim tradition.

“Look,” said Bekka. “The thing about sailing to Tahiti would work a lot better if we knew how many blue-eyed versus brown-eyed people were here. If we all had blue eyes, then we could be sure that the Tahiti plan would work, and some of us could be saved. If some of us had brown eyes, then we could choose a number of people to sail to Tahiti that had a good probability of catching enough of the blue-eyed ones.”

“We can wish all we want,” said Enuli, “but if we explicitly knew how many people had blue versus brown eyes, we’d all have to kill ourselves right now.”

“What about probabilistic knowledge?” I asked. “In theory, we could construct a system that would allow us to have > 99.99% probability what color our eyes were without being sure.”

“That’s stupid,” Enuli said, at precisely the same time Calkas said “That’s brilliant!” He went on: “Look, just between the five of us, everybody else back at the village has blue eyes, right?”

We nodded. It was nerve-wracking to hear it mentioned so casually, just like that, but as far as I could tell it didn’t break any taboos.

“So,” said Calkas, “We know that, of the island population, at least 995 of the 1000 of us have blue eyes. Oh, and since nobody committed suicide last night, we know that at least three of the five of us have blue eyes, so that’s 998 out of 1000. Just probabilistically, by Laplace’s Law of Succession and the like, we can estimate a >99% chance that we ourselves have blue eyes. Nothing I’m saying is taboo. It’s nothing that the priests don’t know themselves. But none of them have killed themselves yet. So without revealing any information about the eye color composition of the current group, I think it’s reasonable to make a first assumption that all of us have blue eyes.”

“I’m really creeped out at you talking like this,” said Daho. I saw goosebumps on his arms.

“I do not believe that the same Volcano God who has endowed us with reason and intellect could have intended us to forego their use,” said Calkas. “Let’s assume we all have blue eyes. In that case, the Tahiti plan is still on.”

“Waaiiiiit a second – ” Bekka objected. “If probabilistic knowledge of eye color doesn’t count, then no information can count. After all, there’s always a chance that the delicious sailor could have been lying. So when he said at least one of us had blue eyes, all we know is that there’s a high *probability* that at least one of us has blue eyes.”

“Yes!” said Daho. “I’ve been reading this book that washed ashore from a shipwrecked galleon. Off in Europe, there is this tribe called the Jews. Their holy book says that illegitimate children should be shunned by the congregation. Their leaders thought this was unfair, but they weren’t able to contradict the holy book. So instead they declared that sure, illegitimate children should be shunned, but only if they were *sure* they were really illegitimate. Then they declared that no amount of evidence would ever suffice to convince them of that. There was always a possibility that the woman had secretly had sex with her husband nine months before the birth and was simply lying about it. Or, if apparently unmarried, that she had secretly married someone. They decided that it was permissible to err on the side of caution, and from that perspective nobody was sufficiently certainly illegitimate to need shunning. We could do the same thing here.”

“Yes!” I said. “That is, even if we looked at our reflection and saw our eye color directly, it might be that a deceiving demon is altering all of our experience – ”

“No no NO,” said Calkas. “That’s not right. The Tablets of Enku say that *because* people must not know their own eye color, we are forbidden to talk about the matter. So the law strongly implies that hearing someone tell us our eye color would count as proof of that eye color. The exact probability has nothing to do with it. It’s the method by which we gain the information.”

“That’s stupid,” Bekka protested.

“That’s the law,” said Calkas.

“Let’s do the Tahiti plan, then,” I said. I gathered five stones from the floor of the lodge. Two white, three black. “White stones stay. Black stones go to Tahiti. Close your eyes and don’t look.”

Bekka, Calkas, Daho, and Enuli all took a stone from my hand. I looked at the one that was left. It was black. Then I looked around the lodge. Calkas and Enuli were smiling, white stones in their hands. Bekka and Daho, not so much. Daho whined, looked at me pleadingly.

“No,” I said. “It’s decided. The three of us will head off tonight.”

Calkas and Enuli tried to be respectful, to hide their glee and relief.

“You guys will tell our families what happened?

They nodded gravely.

We began packing our things.

* * *

The dark clouds frustrated any hope of moonlight as Bekka, Daho and I set off to the nearby cove where two canoes lay hidden beneath the overhanging rocks. The rain soaked our clothes the second we crossed the doorway. The wind lashed at our faces. We could barely hear ourselves talk. This was a *bad* storm.

“How are we going to make it to the canoes in this weather?!” Bekka shouted at me, grabbing my arm. I just squeezed her hand. Daho might have said something, might not have. I couldn’t tell.Between the mud and the rain and the darkness it took us two

hours to travel less than a mile. The canoes were where we had left them a few days before. The rocks gave us brief shelter from the pelting rain.

“This is suicide!” Daho said, once we could hear each other again. “There’s no way we can make it to Tahiti in this! We won’t even be able to make it a full mile out!” Bekka nodded.

“Yes,” I said. I’d kind of known it, the whole way down to the cove, but now I was sure. “Yes. This is suicide. But we’ve got to do it If we don’t kill ourselves tonight, then we’ve just got to go back to the lodge. And then we’ll all end up killing ourselves anyway. And Calkas and Enuli will die too.”

“No!” said Daho. “We go back, we tell them that we can’t make it to Tahiti. Then we let *them* decide if we need to commit suicide or not. And if they say yes, we draw the stones again. Four black, one white. One chance to live.”

“We already drew the stones,” I said. “Fair is fair.”

“Fair is fair?” Bekka cried. “We drew stones to go to Tahiti. We didn’t draw stones to commit suicide. If the stone drawing obliged us to commit suicide, they should have said so, and then maybe we would have spent more time thinking about other options. Why do we have to die? Why can’t the other ones die? Why not Enuli, with that stupid bone in her hair? I hate her so much! Ahuja, you can’t just let me die like this!”

That hurt. I was willing to sacrifice my life, if that was what it took. But Bekka was right. To just toss ourselves out to sea and let her drown beneath those waves would break the whole point of our betrothal bond.

“Well, I – ”

“Ahuja,” said Bekka. “I think I’m pregnant.”

“What?”

“I missed my last period. And I got sick this morning, even though I didn’t eat any extra sparkroot. I think I’m pregnant. I don’t want to die. We need to save me. To save the baby.”

I looked at the horrible waves, watched them pelt the shore. A few moments in that, and there was no doubt we would capsize and die.

“Okay,” I said. “New plan. The three of us go back. We tell them that we couldn’t get to Tahiti. They point out that another night has passed. Now four of us have to die. The three of us vote for everybody except Bekka dying. It’s 3-2, we win. The rest of us die, and Bekka goes back to the village and the baby lives.”

“Hold on,” said Daho. “I’m supposed to vote for me to die and Bekka to live? What do I get out of this deal?”

The Tablets of Enku say one man must not kill another. So I didn’t.

“You get an extra day!” I snapped. “One extra day of life for saving my betrothed and unborn child. Because we’re not going back unless you agree to this. It’s either die now, or die tomorrow night. And a lot of things can happen in a day.”

“Like what?”

“Like I don’t know. We might think of some clever way out. Enku the Lawgiver might return from the dead and change the rules. Whatever. It’s a better deal than you’ll get if you throw yourself into that water.”

Daho glared at me, then weighed his options. “Okay,” he snapped. “I’ll vote for Bekka. But you had better be thinking *really* hard about those clever ways out.”

Day Three

“So,” said Calkas the next morning. “I guess all of us are still here.” He didn’t really sound surprised.

I explained what had happened the night before.

“It’s simple,” Calkas declared. “The Volcano God is punishing us. He’s saying that it’s wrong of us to try to escape his judgment by going to Tahiti. That’s why he sent the storm. He wants us all to stay here until the bitter end and then, if we have to, we die together.”

“No!” I protested. “That’s not it at all! The taboo doesn’t say we all have to die. It just says we all have to die if we figure out what our eye color is! If some of us kill ourselves, we can prevent that from happening!”

“The Volcano God loathes the needless taking of life,” said Calkas. “And he loathes his people traveling to other lands, where the sparkroot never grows and the taboos are violated every day. That’s what he’s trying to tell us. He’s trying to close off our options, so that we stay pure and our souls don’t have to burn in his caldera. You know, like Enuli’s will.” He shot her a poison glance.

“My name is – ” she started.

“I don’t think that’s it at all,” I said. “I say the four of us sacrifice ourselves to save Bekka.”

“You *would* say that, as her betrothed,” said Enuli.

“Well yes,” I said. “Yes, I would. Forgive me for not wanting the love of my life to die for a stupid reason. Maybe I should just throw myself in the caldera right now. And she’s carrying an unborn child? Did you miss that part?”

“People, people,” said Calkas. “Peace! We’re all on the same side here.”

“No we’re not,” I said. “So let’s vote. Everyone in favor of saving Bekka, say aye.”

“And everyone in favor of not sacrificing anyone to the waves, and letting the Volcano God’s will be done, say nay.” Calkas added.

“Aye,” I said.

“Aye,” said Bekka.

“Nay,” said Calkas.

“Nay,” said Enuli.

“Nay,” said Daho.

“What?!” I protested.

“Nay,” Daho repeated.

“But you said – ” I told him.

“You promised me one extra day,” Daho said. “Think about it. Calkas is promising me two.”

“No!” I protested. “You can’t do this! Seriously, I’ll kill you guys if I have to!”

“Then your soul will burn in the caldera forever,” said Calkas. “And it still won’t help your betrothed or your child.”

“You can’t do this,” I repeated, softly, more of a mutter.

“We can, Ahuja” said Calkas.

I slumped back into my room, defeated.

Day Four

I gave them the traditional morning greeting. “So, I guess we’re all still here.”

We were. It was our last day. We now had enough information to prove, beyond a shadow of a doubt, that all of us had blue eyes. At midnight, we would all have to commit suicide.

“You know what?” said Enuli. “I’ve always wanted to say this. ALL OF YOU GUYS HAVE BLUE EYES! DEAL WITH IT!”

We nodded. “You have blue eyes too, Enuli,” said Daho. It didn’t matter at this point.

“Wait,” said Bekka. “No! I’ve got it! Heterochromia!”

“Hetero-what?” I asked.

“Heterochromia iridum. It’s a very rare condition where someone has two eyes of two different colors. If one of us has heterochromia iridum, then we can’t prove anything at all! The sailor just said that he saw someone with blue eyes. He didn’t say how *many* blue eyes.”

“That’s stupid, Bekka,” Enuli protested. “He said blue eyes, plural. If somebody just had one blue eye, obviously he would have remarked on that first. Something like ‘this is the only island I’ve been to where people’s eyes have different colors.'”

“No,” said Bekka. “Because maybe all of us have blue eyes, except one person who has heterochromia iridum, and he noticed the other four people, but he didn’t look closely enough to notice the heterochromia iridum in the fifth.”

“Enuli just said,” said Calkas, “that we all have blue eyes.”

“But she didn’t say how many!”

“But,” said Calkas, “if one of us actually had heterochromia iridum, don’t you think somebody would have thought to mention it before the fifth day?”

“Doesn’t matter!” Bekka insisted. “It’s just probabilistic certainty.”

“It doesn’t work that way,” said Calkas. He put an arm on her shoulder. She angrily swatted it off. “Who even decides these things!” she asked. “Why is it wrong to know your own eye color?”

“The eye is the organ that sees,” said Calkas. “It’s how we know what things look like. If the eye knew what it itself looked like, it would be an infinite cycle, the eye seeing the eye seeing the eye seeing the eye and so on. Like dividing by zero. It’s an abomination. That’s why the Volcano God, in his infinite wisdom, said that it must not be.”

“Well, I know my eyes are blue,” said Bekka. “And I don’t feel like I’m stuck in an infinite loop, or like I’m an abomination.”

“That’s because,” Calkas said patiently, “the Volcano God, in his infinite mercy, has given us one day to settle our worldly affairs. But at midnight tonight, we all have to kill ourselves. That’s the rule.”

Bekka cried in my arms. I glared at Calkas. He shrugged. Daho and Enuli went off together – I guess they figured if it was their last day in the world, they might as well have some fun – and I took Bekka back to our room.

* * *

“Listen,” I said. “I’m not going to do it.”

“What?” she asked. She stopped crying immediately.

“I’m not going to do it. And you don’t have to do it either. You should have your baby, and he should have a mother and father. We can wait here. The others will kill themselves. Then we’ll go back to the village on our own and say that the rest of them died in the storm.”

“But – aren’t you worried about the Volcano God burning our souls in his caldera forever?”

“To be honest, I never really paid much attention in Volcano Church. I – I guess we’ll see what happens later on, when we die. The important thing is that we can have our child, and he can grow up with us.”

“I love you,” said Bekka.

“I know,” I said.

“I know you know,” she said. “But I didn’t know that you knew I knew you knew. And now I do.”

“I love you too,” I said.

“I know,” she said.

“I know you know,” I said. I kissed her. “I love you and your beautiful blue eyes.”

The storm darkened from gray to black as the hidden sun passed below the horizon.

Day Five

“So,” I said when the other four had woken up, “I guess all of us are atheists.”

“Yeah,” said Daho.

“The world is empty and void of light and meaning,” said Enuli. “It’s the most Goth thing of all.”

Calkas sighed. “I was hoping all of you would kill yourselves,” he said, “and then I could go home, and my father the high priest would never have to know what happened. I’m sorry for pushing the rest of you. It’s just that – if I looked lax, even for a second, he would have suspected, and then I would have been in so much trouble that an eternity in the Volcano God’s caldera would look pretty good compared to what would happen when I got back home.”

“I think,” said Bekka, “that I realized it the first time I ate the sparkroot. Before I’d even finished swallowing it, I was like, wait a second, volcanoes are probably just geologic phenomenon caused by an upwelling of the magma in the Earth’s mantle. And human life probably evolved from primitive replicators. It makes a lot more sense than some spirit creating all life and then retreating to a dormant volcano on some random island in the middle of the nowhere.”

“This is great,” said Bekka. “Now even if it’s a Mersenne prime day I can eat as many green vegetables as I want!”

“You know Mersenne prime days only come like once every couple of centuries, right?” I asked her.

“I know. It’s just the principle of the thing.”

“We can’t tell any of the others,” Daho insisted. “They’d throw us into the volcano.”

“You think?” I said. “Calkas was saying before that 99% of us had blue eyes, so probably we all had blue eyes. Well, think about it. The five of us are a pretty random sample of the island population, and all five of us are atheist. That means there’s probably a lot more. Maybe everybody’s atheist.”

“Everybody?”

“Well, I thought Calkas was like the most religious of anybody I knew. And here we are.”

“I told you, I was just trying to behave so that I didn’t get in trouble with my father.”

“What if everyone’s doing that? Nobody wants to get in trouble by admitting they don’t believe, because if anybody else found out, they’d get thrown into the volcano. So we all just put on a mask for everybody else.”

“I figured Ahuja was atheist,” said Bekka.

“You did?!” I asked her.

“Yeah. It was the little things. When we were hanging out. Sometimes you’d forget some rituals. And then you’d always shoot these guilty glances at me, like you were trying to see if I’d noticed. I thought it was cute.”

“Why didn’t you tell me?”

“You’d have freaked out. You’d have had to angrily deny it. Unless you knew I was atheist. But I couldn’t have told you that, because if I did then you might feel like you had to throw *me* in the volcano to keep up appearances.”

“Bekka!” I said. “You know I would never – ”

“I kind of suspected Calkas was atheist,” said Daho. “He got so worked up about some of those little points of law. It had to be overcompensating.”

“Hold on hold on hold on!” said Calkas. “So basically, we were all atheists. We all knew we were all atheists. We just didn’t know that we knew that we were all atheists. This is hurting my brain. I think I’m going to need more sparkroot.”

A sunbeam peeked through the wall of the lodge.

“Storm’s over!” Bekka shouted gleefully. “Time to go back home!” We gathered our things and went outside. The sudden sunlight felt crisp and warm upon my skin.

“So,” said Daho, “we don’t mention anything about the sailor to anyone else back at the village?”

“Are you kidding?” said Calkas. “I say we stand in the middle of town square, announce everybody’s eye colors, and then suggest that maybe they don’t believe in the Volcano God as much as they thought. See what happens.”

“YOU ALL HAVE BLUE EYES!” Enuli shouted at the jungle around us. “DEAL WITH IT!” We laughed.

“By the way,” I told Enuli. “While we’re airing out things that everybody knows in order to make them common knowledge, that bone in your hair looks ridiculous.”

“He’s right,” Daho told her.

“It really does,” Calkas agreed.

“You watch out,” said Enuli. “Now that we don’t have to reserve the sparkroot for interpreting taboos, I’m going to invent a death ray. Then you’ll be sorry.”

“Hey,” said Daho, “that sounds pretty cool. And I can invent a giant aerial dreadnaught to mount it on, and together we can take over Europe and maybe the next sailor who gets shipwrecked on our island will be a little less condescending.”

“Ha!” said Enuli. “That would be so Goth.”

Sun on our backs, we took the winding road into the village.