Slate Star Codex
SELF-RECOMMENDING!

MEDITATIONS ON MOLOCH

POSTED ON JULY 30, 2014 BY SCOTT ALEXANDER

[Also available as podcast here]

I.

Allen Ginsberg's famous poem on 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 capit* 0 comments since 2016-01-31 18:
> *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

one dollar bill. From a god's-eye-view, clearly people should not pay $1 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 b 0 comments since 2016-01-31 18: 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 oth 0 comments since 2016-01-31 18: 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

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 working 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 familia 0 comments since 2016-01-31 18: 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 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!)

0 comments since 2016-01-31 18:

…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 crea[0 comments since 2016-01-31 18:] 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

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 crushing soulless jailhouse

(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 b 0 comments since 2016-01-31 18:

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

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 ra(0 comments since 2016-01-31 18: 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* 0 comments since 2016-01-31 18:
> *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

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 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