

Cave scratchings

**Essays and other writings- leftwing, philosophical and otherwise
2018-2022**

By Philosophy Bear

Table of contents

Table of contents	2
Preface	6
The seven best liked essays	8
This is it: The medium term future in AI and how the left-wing should respond	9
The Ballad of Reading Gaol as a rejection of all law and politics	32
A brief reply to Scott Aaronson's- "We Are the God of the Gaps"	41
Everywhere you go, you always take the weather with you	44
Granting textual authority to overcome textual tyranny	56
The Mystery of Love	75
The problem of simulator evil	85
Social commentary	104
Movements are always a distorted lens on the ideas they embody	105
The greatest potential for censorship is in the algorithm, not in the bans	111
Yvne: The forgotten opposite of envy	121
The paranoid style in petit-bourgeois politics	123
Twitter is a reverse panopticon: The internal agent	125
Government Watch: An idea for an NGO	128
Notes on the Tyranny of Meritocracy by Michael Sandel	130
Why don't people often try to earnestly persuade others of their political views on the internet?	138
The mysteries of welfare economics	146
Two senses of caring	155
Relative income effects can be a lot of different things, not just envy	159
Why we can't (usefully) dismiss concern about the income of others as envious	162
Chesterton's fence and thinking using sayings	165
Musings on the self-organization of authoritarianism	169

Yearning	173
Oh death, where is the antidote for thy sting? Or: Prolegomena to a new philosophy of the Common Task	174
Perspectival fever: On being shot through with philosophical desire	184
On Klutzes	190
Existential tragedies—a partial list of the fundamental complaints of being a person.	194
Try to always be kind because you never know when you're incompetent	196
Brief Reflections	199
The questions that haunt me at 3 in the morning	202
Autopsy on a dream	210
Confessions: A psychological and intellectual autobiography	213
Artificial intelligence	229
Why I don't think identity verification will save us from the coming bot-swarm	231
GPT-3 is -right now- already more than capable of enabling student plagiarism	235
Recent advances in Natural Language Processing—Some Woolly speculations	244
Regarding Blake Lemoine's claim that LaMDA is 'sentient', he might be right (sorta), but perhaps not for the reasons he thinks	250
Against John Searle, Gary Marcus, the Chinese Room thought experiment and its world	255
The AI Control Problem in a wider intellectual context	263
Fiction, Criticism & Similar	275
Just read the Damn Book of All Hours- my (losing) entry for Scott Alexander's book review contest	276
The concept of cringe is cringe	291
The Culture Novels and the deaestheticisation of politics	294
Hypotheticals	297
The Romance of Quantum Archaeology	303
Just three dumb guy thoughts I had, plus a dumb guy idea for a novel	312
A novel I wanted to write: The Honesty Contagion	318

300 arguments, a commentary	320
The adolescent v childish theory of aesthetics	330
For the Left	331
Thinking about political persuasion from a left-Wing point of view	332
Where Philosophy Bear stands on Transgender issues	344
Miscellaneous thoughts on the right wing of politics	348
The big problem with the left	360
A very brief argument for materialism about power	363
Economic justice and climate justice are not metaphors: A response to Justice Creep by Scott Alexander	365
Brief observation on the political economy of the welfare state	374
The nasty side of my leftism	375
Everything is negotiable on the right (and left)	379
A katana, an iron bar, and prison	381
GDP undervalues government services	387
I don't know how to tell you that politics is about murder	390
For communism and against foreclosure on the future	393
The egalitarian past (and future?) of politics	397
Mistaken Identity and misunderstood interests: Haider and identity politics	399
Seeing like a communist, an introduction to Marxism for an intelligent person who hasn't read on it previously	405
Of fake fire-starters and real panic-mongers	416
Philosophy	418
Meeting Nietzsche at the limits of rationality and the limits of Analytic Philosophy 419	
Four parts of belief	429
A sketch of a layered solution to the interpersonal comparison problem	433
The Paradox of the Crowd	448
Why I left philosophy	451

How to do things to words: mapping a post-analytic philosophy of concepts and intuitions	459
Against Libertarian Criticisms of Redistribution	472
Through-going subjective Bayesianism as a solution to the problem of skepticism	479
I'm not particularly worried about confirmation bias	481
Utilitarianism is an egalitarianism	484
An agnostic philosopher considers God, the limits of reason, and universalism	499
An introduction to Jesus of Nazareth considered as an ancient moral philosopher	506
Psychology	517
Harm OCD, a brief introduction	518
Obsessive Compulsive Disorder and the origins of religion	524
A final quotation:	531
Lessons I squeezed from a lifelong severe mental illness	532
OCD, mental illness and "cancel culture"	536
Results of the survey on Dark Secrets, part one	544
Poems	558
Deadwater	559
Afterword to Deadwater	570
Upon seeing a rose encased in ice	572
Untitled	573
Dying in the rain	574
Narrow escape	575
Glide your gaze across blank walls	576
Dew	578
Conversation in a garden about a trampling	579
Futility	580
Moonrise	581
Untitled (slam poetry entry)	582
By a lake with friends	583

Break silence	584
The Ballad of Tim	585
I miss you - a lyric	587
On reading your own writing	589
Stuff that I was on the fence about including	590
Why this section?	591
Scientific realism and the moral uses of science	592
Carving up the philosophical terrain around personal identity a little differently	594
Should you care about that issue?	596
Conservation of moral status under misfortune	597
Some stolen ideas on dynamic memetics	600
Money and the sceptic	604
On the perils of contrasting niceness with kindness	607
What form of ethics is the least subject to rationalising what you wanted to do anyway?	611
Paradox of the book and the robot	613
Two reasons why it's easier to find an audience for writing on the basis of different experiences than novel ideas	614
The paradox of high expectations: The more you demand, the less you get	616
New thought experiments for the backyard metaphysician to try at home	619
On critical social-technological points	622
Final bonus	625

Preface

I've always been obsessed with death. Part of that obsession is a sense that, at any given moment, as far as is possible, I should have my affairs in order. Part of that is a sense that I should have my writings collated and neatly organized. Do not ask me how that makes sense, it just fits together in my head. That's why I publish these little collections of my writing.

Twice already I've published a collection of my best writings. I've found it to be a delightful process that shows me that I've done more than I realize. But what does this collection offer the reader? It's hard to say something general about such a collection, but I believe that some of my best writing has been done recently. It's a much more substantial volume- lengthwise and intellectually- than my previous two publications. Also, depending on where you get it, it's free or cheap.

Fortunately I've become a better writer with time, Unfortunately, I haven't got enough time to rewrite all my old work. Nonetheless, I've included old work when I think it says something important, even if it's rough.

It would upset me if this book raised anyone's blood pressure or wasted anyone's money. Thus I say, there are three sorts of readers who might be infuriated by this book and who might be best advised to avoid it:

1. People who hate a style of writing built on *musings dilettantism*. Throwing ideas to see whether they stick.
2. People who get really angry at *left and far left writing*.
3. People got the kindle version not because they wanted to support me, but because they were unaware that my Substack has a free PDF version- only to later find out about the free version.

If you are any of these types *you have been warned*. I'd rather not take your money or your time if it's going to make you unhappy.

All that said, I do my best to treat writing as like hosting. You're my guest. Given that you're mute here, I need to be extra attentive to your needs. Treat this book how you like. Read it in order, read it by section relevant to you or jump straight to the essay titles that grab you.

Dedications. Thanks to mum and dad, to Kieran Latty, to numerous friends who I've bounced ideas off over the years, to my readers, and to my teachers.

P.S. this is cringe but... I can sing and write lyrics. I can neither play nor write instrumental music. I've always wanted to work on making songs. If you're interested in collaborating, get in contact sometime.

The seven best liked essays

[Note: this is based on my judgments as to which of my essays have been and will be most popular, not *my personal assessment of how good they are*. Most of the stuff that I consider my best work isn't in this section. Indeed, some of the stuff that I consider my best work is in the "stuff I was on the fence about including" section!]

This is it: The medium term future in AI and how the left-wing should respond

We need to start treating AI development, and its potential impact on the possibility of a humane world, as seriously as we treat climate change. I'm not even talking about existential risk or far-flung, distantly possible applications. I am talking about things that are coming in the next half-decade. I'm talking about stuff that's technically already possible but is still in the implementation phase.

Computers eating us all is scary, and I'm glad there are people working to prevent it, but there are many other ways artificial intelligence can go very wrong very soon. We need to get real about this on the left. AI futures has a reputation as only being of interest to techno-utopians, and techno-dystopians. We need to think much more seriously about how technology will reshape the battlefield of the struggle for a better world.

I don't think that we can stop what's coming in the near future through ludditeism. For one thing, the technologies I talk about here are more or less already 80% developed. So let's have a think about what's coming and how to respond together.

The world I am picturing

Based on current trends the world I'm talking about five years in the future, give or take three years. In the next section, I'll give some evidence for that claim.

In this world, which I'll sometimes call the medium-term future, computers can do a lot and can take on many jobs. However, there is not (at least not yet) an intelligence explosion where recursively self-improving computers make themselves smarter and smarter, seemingly without limit. Direct human involvement is still necessary for many important white-collar processes, but not all.

The two main consequences I foresee from this which are directly relevant to our discussion are:

Massive job loss- at least in the short term- as a variety of mostly middle-class jobs start to go out the window. Physical labor is, perhaps surprisingly for some, relatively late to start shedding jobs.

There will be computers that can produce propaganda- good propaganda- on an industrial scale. Think of the liberal moral panic over the all-powerful Russian bots but actually significant- not just producing gobbledygook for 13 Twitter followers. All making cartoons, memes, ads, podcasts, youtube clips, essays, merchandise, and so on.

Exactly how long this medium-term future will last is not clear to me. It could be one year or twenty between this and the singularity. What happens in this interregnum between our period and true artificial superintelligence (ASI), or to put it more crudely, who is on top of the pile when ASI comes along, could be one of history's most important questions.

The pace of advances in the last few years

In a previous essay, I gave evidence for the claim that language models are getting better at a staggering, even menacing pace. Computers are cracking challenges and datasets that only true AGI was supposed to be able to crack like the Winograd schema, which was intended to be a replacement for the Turing test. However, in this essay, I'm going to make the point with samples, rather than measured trends. If you want the latter, check out my essay here.

If you want more specific detail on the rate of progress in language models read my essay on verbal parity here.

In what follows, I'll give examples of what two models can do. PALM 540B is a language model, while DALLE-2 is an image and language model. I hope you'll agree their progress is staggering.

PALM 540B

Consider PALM 540 B, probably the most sophisticated language model currently publicly known of.

Here's Palm 540B explaining a joke:

Prompt: Explain this joke:

Joke: Did you see that Google just hired an eloquent whale for their TPU team? It showed them how to communicate between two different pods!

Model response:

Prediction: TPUs are a type of computer chip that Google uses for deep learning. A "pod" is a group of TPUs. A "pod" is also a group of whales. The joke is that the whale is able to communicate between two groups of whales, but the speaker is pretending that the whale is able to communicate between two groups of TPUs.

One more joke explanation, because this one was quite impressive:

Input: I tried 10,000 random restarts of my neural network, but I was accused of overfitting. I guess no good deed goes unpunished.

Model output: This joke is a pun. A neural network is a computer program that can learn from data. A "seed" is a number that is used to initialize a random number generator. A "good seed" is a number that produces a good random number generator. The phrase "no good deed goes unpunished" means that if you do something good, you will be punished for it.

Here it is engaging in two logical inference chains, the second one is particularly impressive:

Input: When I found out my grandma was in the hospital I felt a particular color. When someone cut me off in traffic I felt a different color. What is the most likely color I would see if I combined these two colors?

Model Output: The color that I felt when I found out my grandma was in hospital was likely a sad color, like blue. The color that I felt when someone cut me off in traffic was likely an angry color, like red. If I combined these two colors, I would likely see purple. The answer is purple.

Input: Michael is at that really famous museum in France looking at its most famous painting. However, the artist who made this painting just makes Michael think of his favorite cartoon character from his childhood. What was the country of origin of the thing that the cartoon character usually holds in his hand?

Model Output: The most famous painting in the Louvre is the Mona Lisa. The artist who made the Mona Lisa is Leonardo da Vinci. Leonardo da Vinci is the name of the main character in the cartoon Mutant Ninja Turtles. Leonardo da Vinci is from Italy. The thing that Leonarda da Vinci usually holds in his hand is a katana. The country of origin of the katana is Japan. The answer is “Japan”.

DALLE-2

Dalle-2 is a multimedia bot that works at the interface between language and imagery. For example, it can create images from captions. Right now, just as is, DALLE-2 could displace an awful lot of graphic design work.

In the original piece, at this point, I included some impressive creations by DALLE-2. Unfortunately copyright law prevents me from including them in this book, which is to be

published on Kindle. I can promise you though that the images were spectacular responses to very specific, detailed requests for images. Just Google “DALLE-2 Images” and have a look at some of the things it’s made, and the requests it made them in response to. Don’t confuse it with “DALLE-Mini” now called CrAIyon, which is cute and publicly available, but nowhere near as impressive.

Replies to common objections

We still see a lot of arguments over whether or not language models “really” possess certain capabilities, most often, commonsense. What I would say is that numerous ways of measuring commonsense intended to trip up language models have been devised. Pretty much all of them have been overcome by Natural Language Understanding, including many that were intended to be insoluble by anything except “true” AGI. For example- the Winograd schema- over which a big deal was once made.

Tangent- I find it bizarre that this hasn’t been received as more of a blow against the language of thought hypothesis and a blow for latter-day versions of associationism in cognitive science. If I were a language of thought guy like Pinker I’d be feeling pretty worried about now. Maybe I’m just not hooked into the debates. Steven Pinker, by the way, wrote this back in 2019, just before the most impressive achievements of Deep Learning started to bloom. It would be fascinating to trace the connections between his peculiar brand of small c-conservative liberalism, and this colossal misjudgment. While I wouldn’t call myself a blank slate, ML has definitely moved me somewhat in that direction.

He said this in 2019, right before stuff started heating up:

Don’t get distracted by debates over whether or not this is real AI

I would urge everyone involved in these discussions to avoid interminable debates about whether contemporary AI and ML constitute “real” intelligence or “real” understanding- or at least to keep such discussions in their proper place and in perspective. We’ve all heard of the Chinese room thought experiment, but here’s the thing. Suppose you were to put the Chinese room to work as an English/Chinese translator and interpreter its economic value in this role would be not a whit affected by the question of whether it really understands Chinese or not. History is moved by capabilities, by what things can do.

Preliminaries aside, let’s start with a few predictions of what this world is going to be like, and move on to discussing how we can fight for a better future in it.

Prediction 1. there will be a flood of AI-generated propaganda

Very soon it will become even harder for this blog to be seen. AI that can write reasonably insightful blog posts, and tout its own work better than I can are coming. The same is true for memes, novels, audiobooks, etc.

The people who operate these engines of cultural production will in almost all cases not be benign- who with a great deal of money is? They will certainly not be politically or socially neutral. The essays, memes, books, soundbites, and tweets generated will, I’m sure, be very persuasive. Imagine powerful people able to buy unlimited amounts of rhetorical firepower at much cheaper rates and higher quality than what has previously been the going standard- at least for mass-produced media.

They will all seek to advance their interests. The internet will swarm with bots, but unlike the bots of yesteryear, they’ll be charming and compelling. Tech giants won’t don’t just source (and aggressively curate) content, they’ll produce it at a superhuman rate, maybe even directly customized for each user.

They'll be smart about it too. They won't just write essays no one reads like I do. They'll form relationships and mimic friendship. They'll hijack existing discussions almost seamlessly. They'll create communities and subcultures and lonely people, old and young, will seek them out. All under the command of the lucre-holder bought at bargain-basement prices- if current prices are a guide. GPT-3 will write you 16000 words for a dollar. Imagine how much mind-junk a billionaire could buy at that rate?

As for us content producers- only those of us who were famous prior to the AI content bloom will have a chance, and I suppose also those who can market their work as specially connected to their identity and life experiences in a way machines can't match.

Edit: I saw someone responding to this post by saying that social media platforms could avoid this flood of bots by introducing ID vetting to make sure users are human. Some might, but I think that, in a lot of cases, it will be the social media companies employing the bots to make their websites more engaging and content rich. I also think that, in a lot of cases, users will seek out bots themselves.

Under a flood of AI propaganda, people will just stop giving a f* about persuasive material**

In a world where well constructed, entertaining, and persuasive material- books, threads, memes, essays, youtube videos, podcasts, etc. that argue for just about anything is produced at an inhuman rate, people will simply stop giving arguments credit because they are hard to refute or eloquently put. A handful of people might start to emphasize the importance of deferring to experts, but for the most part, the reaction is going to be that people "go back to their gut". People will try to reason through issues on the basis of fundamental values, and core ideas about human nature, rather than the relative merits of arguments. Of course, the bots will get good at manipulating us as we try to avoid them by playing this game too.

In 99% of cases, we're already there! How much does public reason really matter? But the vestigial wings of logic will fall away as the sky gets too tumultuous to fly, crowded with eloquent, learned, and inhumanly tireless sophists. My hunch? Even though most of the time giving reasons is not that effective, the loss of that last tiny little bit of space for it will be more significant than you think. But who knows?

Of course, this won't all be rolled out evenly. There will be late adopters who keep forming their views on the basis of things they see and read that seem to make a good point. Ironically, these people will be regarded as extremely gullible. This process is already underway to some extent- even before the flood of AI propaganda. Many of us have groaned at the earnest centrist who butts in "but I read this study from a think tank that suggested..."

AI, in response to people becoming increasingly jaded by persuasive material, might become more targeted in frightening ways. E.g. forming online friendships under the cover of humanity (or not) and subtly pushing its preferred political angles rather than openly spewing propaganda.

They'll have us talking culture war bullshit non-stop

They say that all predictions of the future are secretly about the present. Well, you've got me here because there's one feature of our contemporary situation that I think will be greatly exacerbated by bots: endless fucking discussion of the culture wars. If you were a sociopathic rich guy with an army of bots who wanted to whip people into a frenzy to cut your taxes or regulate out of existence your competitors- isn't it what you'd go for? Tireless computers will scour the internet for trans children for conservatives to scream about. The very same algorithms will hunt down evidence of people having said a slur 20 years ago or some other drivel for liberals to get mad at. These aspects of AI will intensify (if you can believe it!) the worst features of the current period. We will all be

under the gaze of moralistic engines, with no compunctions against impugning us, with no history or hot buttons to hit back on.

Prediction 2, a lot of people will lose jobs: “Technological unemployment”- temporary or not, it doesn’t necessarily matter

AI is getting better and better at doing human tasks, and so will be put to work. Jobs and tasks centered on writing, reading, drawing, programming, customer conversations (including speaking via phone), and more are all within striking distance.

There is a debate over whether machine learning will cause unemployment in the long run. The best argument it won’t is inductive- in the past new technologies have created more jobs than they have destroyed. The basic argument it will is that induction doesn’t work because no past technology is so ubiquitous as contemporary deep learning in replacing human intelligence. It targets more or less all distinctively human capabilities. Taken to the limiting case we can imagine a situation in which literally any activity can be done cheaper by a machine than a human, where even hiring a human at zero dollars an hour would be less cost-effective than getting a computer to do it, due to sundry employment costs like insurance.

But whether replacement jobs will be found eventually- whether the technology I am interested in affects the equilibrium unemployment level- is somewhat beside the point. There will be big job losses in the short term and that’s the most important thing from the point of view of this essay.

Of horses- an aside on technological unemployment

As an aside, let me go into the long-term aspects of the technological unemployment debate. I've seen the following debate play out a fair few times:

A. Cars replaced horses. Why couldn't humans be replaced by robots?

B. These things are more complex than that. Yes, the number of horses is massively down from a century ago, but in the past few decades, the number of horses is now actually growing! New roles are found for old things, and technology means we can have more of everything.

Personally, I think both sides of the argument are missing the main point here. As above, I predict there will be job loss in the short term and don't know if those jobs will eventually be replaced, but what I do know is this.

Suppose in 1900, the horses of America had decided to go on strike. Their strike would have had considerable heft. Doubtless, the farmers would have had to come to the negotiating table pretty soon to increase hay allocations and get rid of the glue factory. All of American society might have been forced into a great compromise with the horses. On the other hand, suppose horses went on strike now. I doubt many people would care all that much.

Jobs may or may not be replaced, but there are many ancillary questions- for example, could humans become less economically indispensable? These questions have implications in themselves. Attempts to avoid them by appealing to what technological progress has done in the past seem of limited use.

And again! I cannot emphasize enough that Job churn due to Robo-replacement matters, even if new jobs are created. Getting fired is 47 points on the Holmes and Rahe stress scale, a bigger deal (apparently!) than falling pregnant, just slightly less

transformative than marriage or a significant personal illness or injury and not even that different from being imprisoned (63).

Prediction 3: the interaction effect between job loss and propaganda will be very bad.

Consider the combined impact of an unlimited, well-designed propaganda generator combined with massive unemployment

I don't know what a world in which it's very cheap to produce very good propaganda, and also people are rapidly losing their jobs, looks like. I have a few unpleasant guesses.

To spell it out, far-right power seems like a plausible outcome in a world of unemployment, brimming with fecund propaganda machines ready to orchestrate discontent, controlled by rich people.

Prediction 4: The world will throw egg in our faces. Too many predictions are ill-advised

Apollo sends the Erinyes to hunt those who try to usurp the gift of prophecy through thinking hard alone. In a somewhat milder than normal temperament, they just make us look like fucking idiots. This is my attempt at a restrained, modest series of best guesses, and nevertheless, I bet even a lot of what I say here will be wrong or irrelevant.

This essay is mostly about a medium-term AI future. There are a lot of uncertainties around this- how long will this period last? Will it be followed by a singularity? Who will have control over the machines and to what degree will they monopolize this

power? I don't know. I'm trying to stick to a very flat projection, and not create elaborate models based on branching dependencies and assumptions.

I recommend you think about AI futures like this, as far as you can. Focus on very straight-line projections, and think about building power in ways that will help almost regardless of how things turn out.

Let me give one example of a wildcard that could throw out a lot of guesses about what the future will look like. For a long time now I've been fascinated by the possibility of creating a neural lie detector that actually works. A society with such a lie detector would change in ways we can't even begin to anticipate. The criminal justice implications are obvious but are far from the most significant. Politics, business, interpersonal relationships, the very concept of honesty... I have a suspicion that certain advances in machine learning might bring us much closer to such a lie detector, but I can't know. If it happens, things could be very different [or, on the contrary, shockingly unchanged], so I've tried to keep low with my predictions, and only go with very direct implications of things that already halfway exist.

Another wildcard just to make the point. We've all put dirt about ourselves out there on the internet. AI could get very good at hunting it down. That could mean blackmail becomes very important in the future, or it could mean that we all get jaded of scandals, or successfully develop a convention of ignoring scandalous material if it was obviously revealed as part of blackmail. Khe sera sera.

Of Critical-Technological Social Points

In the past I have written about what I call critical social-technological points

CSTP is a technological discovery or implementation after which the existing hierarchies and ruling class of a society is locked in, in the sense that removing them

from power, or even resisting them in any materially important way, becomes much more difficult.

Saying that technology is a CSTP is different from saying it is an inherently authoritarian technology- A CSTP generally only threatens authoritarianism if it is achieved in an already authoritarian society.

General AI, especially artificial superintelligence, seems likely to be a critical-technological-social point. It's the big one. If a small cadre of people have power over the first superintelligences, they will be able to use the enormous might that grants them to cement their rule, maybe out to the heat death of the universe.

How well we deal with this medium-term, pre-super-intelligent AGI scenario I describe might make all the difference. It might determine, so to speak, the values of the ascenders, and which way this CTSP pushes us.

Who is developing the technology at the moment

In the west, ML technology is largely being developed by tech giants owned by rich people and institutional investors (who disproportionately represent rich people). Doubtless, these institutions have links to the national security state and even to partisan political interests, but they are kept tastefully discrete [this is no compliment]. In China, state power is more upfront, though still often mediated through rich people (c.f. Ali Baba). Neither option appeals to me. However, the cash demands of creating these machines- currently up to 17 million dollars just for direct computational costs- and vastly more when other expenses are included- mean that it's difficult for university researchers, let alone amateurs, to do this kind of research. In some ways that might be for the best, we don't want a backyard super-intelligence to kill us all. In other ways, it's less desirable.

The role of workers

For those of us who want a future in which ordinary people matter politically and have their needs met, what comfort? The workers who develop these algorithms are mostly humane and decent, and this kind of complex work and research process might give them opportunities to inject their values into the mix, for example, by slowing down research with directly malicious applications, and by acting as whistleblowers. This is some hope then, at least in the medium-term future, we are talking about, before the algorithms start developing themselves without human involvement.

The symbolic role of workers will also be huge, in the debates over who deserves what. Big tech companies will say “hey, we deserve these mega-profits we’re pulling in, and the social and political power we hold because we created this groundbreaking technology. What better response is there to that except to say “no actually, the workers made it”?

Suggestion 1: Reach out to AI researchers

There’s an old riddle- the riddle of history. Who in our society has A) no real material attachment to the present order and B) the power to change the present order. It’s like a murder mystery- who has both the motive and the means? The general answer to this riddle of history is workers- workers have the power to change the world, and few incentives to keep it as it is.

Around these types of questions, a particular type of worker becomes extra important- AI researchers. AI researchers don’t have monetary skin in the game in the form of equity in a company. They’re not the kind of strange, sociopathic or at least situationally sociopathic creatures that come to own a substantial chunk of something like Alphabet or OpenAI. They don’t share the interests of these companies. They do have power over the process. In my experience, they tend to take the ethics of AI research very seriously- go on r/machinelearning and look at how they respond to topics like facial recognition

or criminality prediction- see for example this thread- they're not all ML researchers in the thread, but they seem pretty representative of the sentiment from what I can tell.

So I suggest reaching out to AI workers. Making connections with ML researchers through various forums- union building, AI ethics research and so on- and making sure there is space for ML researchers to talk about the kinds of models they believe in building, and the kind of world they want their models to contribute to- all seem like a solid idea.

Will the military power of ordinary be increased or decreased- we don't know

There's a dynamic in history where some technological advances give military power to elite people, and some technological advances give power to weak people. The stirrup gave a great deal of power to elites. The gun gave a great deal of power to ordinary people. Combined arms military doctrine took a lot of power away from ordinary people. The technologies of guerilla warfare gave it back. It's useful to think about the medium-term AI world I'm describing from this angle. Will it make it easier or harder for ordinary people to resist the power of states?

The nightmare scenario is a world where technology means that what individuals outside a handful of elites wants doesn't matter. The vast majority of people are left with effectively zero military power. Any power they hold is granted through the (tenuous) generosity of elites. The military itself consists in a series of drones and a few human controllers who are thoroughly bought and paid for. The kind of medium-term future I'm talking about might generate this outcome.

But there's another possibility in which the kinds of medium-term AI I'm talking about actually disperse military power, at least for a little while. I wouldn't have thought it possible until the Ukraine war started, and it became clear that cheap drones might be the future.

Leaving this question aside, here's another interesting question about the capacity for rebellion. Will the concentration of power into drones increase or decrease the psychological capacity of militaries for brutality against their civilian population? I.e., will it increase or decrease the capacity of militaries to put down revolts? I don't know. The easiest argument to make is that it will increase it, because drone operators will be able to filter out the horror of much of what they do. But it might also decrease the capacity of militaries for brutality against their own people. I'll admit my argument is tenuous, but here's why I think this. My understanding is that brutality against civilian populations is often carried out by special forces psychos and similar desensitized individuals. I don't know how suitable drone operators are for that sort of work at least when they identify with the people being killed.

Prediction 5: Journalists and other high ranking white-collar jobs will flip on job losses

It will be funny to see journalists, etc. who have historically told us not to worry about job losses- assured us that they are simply the price of free trade and globalization- suddenly find within their hearts a deep wellspring of compassion for jobs threatened by AI. It will become clear that their own jobs are on the chopping block and their sentiments will shift very quickly. Maybe even economists will fall in this category.

We might see conflict between high-ranking journalists, academics, lawyers, doctors, etc. who feel confident that their jobs will not be automated due to the value of their personal prestige, and lower-ranking members of these professions who will not feel so secure.

All this is happening amidst an awful tangle- Nuclear tensions, low-intensity civil conflict, rising inequality and global warming

We can't make specific predictions about the future, but we can say all this is happening while things are getting ugly. For example, consider changing attitudes towards political violence:

[In the original there was a chart from Statistia demonstrating a rise in the percentage of US adults who said they feel justified to use violence to advance political goals from about 8% for both Democrats and Republicans in 2017, to about 33% for Democrats and 36% for Republicans in 2020. A massive increase, as these things go. I cannot reproduce this due to copyright.]

Regardless of what I think about the abstract question of whether it is ever defensible to use violence to achieve political goals, I do not think this particular rise is healthy.

Meanwhile, we're crossing red line after red line on the climate. Tensions between nuclear powers are ratcheting up between the US and China and between the US and Russia. Tensions remain high between India and Pakistan.

AI, climate, nuclear tension and political polarisation (of a very unedifying sort!) interlock into matrixes of possible catastrophes.

Suggestion 2: Move offline

For a while now there's been, at least on the left, a movement away from the online world, see for example the so-called "Grillpill". If the internet becomes loaded up with bots- bots that operate exactly like humans and have the agenda of whoever is rich enough to pay for them- getting off the internet and organizing IRL will be more important than ever. We should already be doing this for so many reasons, but this is another good reason to add to the pile.

We should also, in any case, and even not just for political reasons, be trying to create real connections between human beings. Hangouts, clubs, unions, reading groups and so on. Loci of people able to talk politics independently of social media algorithms, and the coming bot plague, give us the best chance at autonomous action.

Who exactly is the government might come to matter a great deal

It's hard to be specific about this- I don't have the details-, but it's very plausible that our future right now is extremely path-dependent. Whoever is in charge during the rise of the bots may have an outsized influence on what post-humanity looks like.

I'm not just talking about obvious things like "who is president". I'm also talking nitty-gritty details. Who's on the Joint Chief of Staff. Who's CEO at Alphabet. How much power Musk and other "big personalities" like him have in the process?

What to think of property when property is not necessary? What to think of people when they are not necessary?

Suppose it looks like we're approaching a post-scarcity society fast. What do debates over the value and relative value of human beings look like then?

A lot of authors have written on what a post-scarcity society might be like- a society where human labor is mostly superfluous. Often these authors are most concerned about the fates of the technologically unemployed- will they be treated as surplus population, fit for marginal survival at best, elimination at worse- or will they be treated well through initiatives like UBI.

But artificial intelligence raises another question, no less important. If the special privileges of the rich no longer serve an economic function from the point of view of society as a whole, if there is no longer an argument to be made that their special

entrepreneurial judgment is needed to enrich us all, will they, nonetheless, be kept in conditions of extraordinary power and prestige?

I suppose the answers to these dueling questions are logically independent. A society that maintained a decent standard of living for everyone, but lifted a few oligarchs to glory, is perfectly conceivable. Even the inverse- a society that threw down the oligarchs, but also used post-scarcity as an opportunity to purge itself of “undesirable” elements whose labor was no longer a reason to keep them around (think racist societies with migrant workers) is possible.

Still, I can't help but think these questions reflect on each other. Will we throw down the strong, or keep them out of a residual capitalist morality, even as the economic structure of capitalism becomes increasingly irrelevant? Will we raise up the economically superfluous, or will we leave them in the dust, reasoning that since they hadn't taken (possibly non-existent) means of lifting themselves out of squalor, they deserved their fate? The connecting thread is capitalist morality. Capitalist morality gives us the sense that people deserve their economic outcome, which reflects on their virtue. It draws our attention away from the role of contingent aspects of people's regulatory, social, and economic environment, the distribution of property and wealth prior to their birth, random talents given by genetics and upbringing, and to top it all off, a healthy serving of luck (think of the distribution of illness alone!) The world is a shambles, but I preach to the choir.

If the coming of superabundance breaks capitalist morality, so much the better for the poor, so much the worse for the rich. If the coming of superabundance does not lead to the break up of capitalist morality, then I worry for the future of the poor (which will be most of us). The world has always hated poor people, but wait and see what happens if the world no longer needs them. Worst-case scenario think South American style classicide death squads and the power of the rich locked in.

Suggestion 3: attack the legitimacy of property

There's one relatively obvious upshot for strategy in the arena of ideas. There have always been two ideological pillars upholding capitalism 1. The contention is that it is necessary to keep the economic system ticking along, to put bread on all our tables, and so on. 2. The contention that capitalism is just.

In a world in which capitalism serves a diminishing economic role, because it is less and less necessary to compel human labor to work, and performing economic planning through artificial intelligence rather than markets is increasingly possible, one of these pillars falls. Will the other pillar be able to remain up on its own? I'm not sure. What I am sure of is that a serious discussion about property, about the ways in which property has been held to both justify, and be justified by capitalism, is necessary.

We need to start making the argument now that there is no inherent justice to the distribution of property, and the concentration of ownership of productive assets into a small of hands that we observe has no correlation with underlying desert. Jeff Bezos does not deserve to be making over six hundred thousand times more an hour than many workers in the United States.

If someone could find a way to make Murphy and Nagel's *The Myth of Property* into widely accessible agitprop, that would be great. Regardless, worth pulling out a copy.

Of course, it's all good and well to try and map out a hypothetical argument at the end of history, but who will be the judges of who wins that argument, in theory, and in practice? Who will have the power to decide justice? Unknowable, but best to start making the argument now.

Reimagining arguments over socialism

Let's define socialism as the position that production should be ordered according to a conception of social welfare (whether informal or a formal social-welfare function), rather than ordered by exchange. Capitalism is the position that production should be ordered by exchange.

Something that I think most political philosophy misses about our political feelings is that they are mostly vectors rather than points in the space of possible political philosophies. What matters is not so much my ultimate preferred society as the direction I'm inclined to want to move things.

Define vector socialism as the position that we ought to be moving in the direction of production and allocation for the purposes of a conception of social welfare- giving these considerations increasing power over what is made and who gets it. Define vector capitalism as the position that we ought to be moving further in the direction of production and allocation for the purposes of exchange.

Another useful concept. Define minimal socialism as the view that, whatever is actually possible, it would be at least be ideal if production and distribution were performed to directly meet human needs, rather than for exchange.

I find myself both a vector socialist and a minimal socialist, the question of whether I'm a socialist simpliciter is somewhat up in the air. I've defined all these terms for a somewhat paradoxical reason- to point out that, under conditions of increasing post-scarcity, the distinctions cease to matter- they all sort of collapse into each other. Technical barriers and feasibility gaps between the preferred and the possible no longer matter. All that remains is a kind of ethics quiz at the end of time. Do we want inequality between persons or not?

The trouble is we don't know who will get to answer the question, but we can do our best.

The Ballad of Reading Gaol as a rejection of all law and politics

“For he who lives more lives than one
More deaths than one must die.”

-*The Ballad of Reading Gaol*, Section III

“They think a murderer’s heart would taint
Each simple seed they sow.
It is not true! God’s kindly earth
Is kindlier than men know,
And the red rose would but blow more red,
The white rose whiter blow.”

-*The Ballad of Reading Gaol*, Section IV

1. Stones on the Mount of Olives

When I was young a number of horrific experiences convinced me that I could either choose to be wholly on the side of humanity—all of humanity—or a misanthrope. I chose the first option, although I fall short constantly. Trying to explain how that commitment to being on the side of humanity works on the level of feeling—to show how certain ideas are emotionally and aesthetically coherent with each other in order to create a harmony in how I feel about humans in general, is what led me to write this essay.

We’ll get to “The Ballad of Reading Gaol” soon, as promised in the title, but before we do I want to take a detour through the Gospel of John. (Don’t worry—I’m an agnostic and this isn’t going to turn into a religious essay.)

One of the most famous passages in the New Testament is the story of the woman taken in adultery. You may remember it as the story with the line: “Let he who is without sin

cast the first stone”. It’s in the Gospel of John:

“[...] Jesus went to the Mount of Olives. Early in the morning he came again to the temple. All the people came to him and he sat down and began to teach them. The scribes and the Pharisees brought a woman who had been caught in adultery; and making her stand before all of them, they said to him, ‘Teacher, this woman was caught in the very act of committing adultery. Now in the law Moses commanded us to stone such women. Now what do you say?’ They said this to test him, so that they might have some charge to bring against him. Jesus bent down and wrote with his finger on the ground. When they kept on questioning him, he straightened up and said to them, ‘Let anyone among you who is without sin be the first to throw a stone at her.’ And once again he bent down and wrote on the ground. When they heard it, they went away, one by one, beginning with the elders; and Jesus was left alone with the woman standing before him. Jesus straightened up and said to her, ‘Woman, where are they? Has no one condemned you?’ She said, ‘No one, sir.’ And Jesus said, ‘Neither do I condemn you. Go your way, and from now on do not sin again’””

Now my friend Karl Hand, biblical scholar extraordinaire, assures me of two things. Firstly, there is almost no doubt that this passage is a later addition, written by another author. Secondly, among the relatively small number of scholars who defend the authenticity of this passage, most are conservatives. However, in my research, I found that, while evangelical and fundamentalist Christians generally defend the whole of the bible, on the grounds that God would not let his word be polluted with error, there is a small grouping of far-right cranks who argue that this passage is, unlike the rest of the Bible, inauthentic. The, uh, always interesting source Conservapedia has it:

“Historians and scholars agree that the story of Jesus and the woman caught in adultery is not authentic and was added decades later to the Gospel of John by scribes. The story was almost certainly added for the purpose of Democrat ideology: if no one who has sinned should cast the first stone, then the message is that no one should punish or even

criticize sinners. It is also clear from the writing style that this story was added later.”

It is most curious, surely, that the very same people who have defended the literal accuracy of the Bible, even to the extent of claiming the world is 6000 years old, are suddenly astute textual critics when it comes to this passage? How overwhelmingly threatening it must be, to be the sole portion distressing enough to move these arch-conservatives away from the doctrine of biblical inerrancy.

2. A threat to society as such

The reason why the strange conservatives at Conservapedia are keen to disavow this, and only this passage is that it proposes, more or less explicitly, that because we all share in the same sinful nature, none of us has the right to punish another. Such a perspective, however impractical it may be, is a conceptual threat to all systems of authority, laws, hierarchy, and ultimately even to organized society. Nonetheless, I think it's one of the best wishes anyone has ever made.

The two oldest functions of government are criminal punishment and defense of territory. This last category might even be seen as a special case of punishment—deterrence through the use of incentives. It's often said that the state is defined by a monopoly on violence; well, the most fundamental form of that violence for the state is punishment. This story of a woman, her accusers, and God become flesh cuts against the very heart of government, conventional morality and capitalism. It is, in the purest, most glorious, and sadly most impractical sense, anarchist.

The same radical message appears in many places, but few as eloquent as “The Ballad of Reading Gaol” almost two thousand years later—but here it comes with a twist.

3. Biographical note

In 1895, Oscar Wilde was sentenced to two years of hard labor for “gross indecency with another man”. He spent much of his sentence in Reading Gaol.

While at Reading Gaol he watched, appalled, as Charles Thomas Wooldridge was executed for the crime of slitting his wife’s throat. Oscar Wilde was a humanitarian, an anarchist, a socialist, and a man who never softened to the world’s cruelties. The idea of executing anyone was truly indecent to him, and he saw the hypocrisy of a violent society punishing violence.

After being released from prison he wrote “The Ballad of Reading Gaol.”

4. The poetic argument

“The Ballad of Reading Gaol” is a poem, and therefore its content cannot be distilled into a list of “points”. As Harold Bloom once said, the meaning of a poem could only be another poem. Yet there are clear themes which, however superficial it may be to do so, we can grab and isolate.

Where “The Ballad of Reading Gaol” differs from the story in John of the woman taken in adultery is that it proposes two reasons why punishment is fundamentally indecent. These reasons are in tension with each other, but not, I think, ultimately contradictory. The first reason is that we are all fundamentally sinful in nature, so whoever performs the punishment is implicitly claiming to be fundamentally different from the punished in a way which just isn’t true. This reasoning can be found in the story of the woman taken in adultery.

The second reason it gives isn’t so obviously present in that biblical story. People are

noble and beautiful, and whatever their flaws, don't deserve the dehumanization, agony and humiliation that comes with punishment, at least as it is practiced in our society.

Describing the prisoners coming out after the morning of the hanging:

“And down the iron stair we tramped,
Each from his separate Hell.
Out into God's sweet air we went,
But not in wonted way,
For this man's face was white with fear,
And that man's face was grey,
And I never saw sad men who looked
So wistfully at the day.
I never saw sad men who looked
With such a wistful eye
Upon that little tent of blue
We prisoners called the sky,
And at every careless cloud that passed
In happy freedom by.”

Or:

“They hanged him as a beast is hanged:
They did not even toll
A requiem that might have brought
Rest to his startled soul,
But hurriedly they took him out,
And hid him in a hole.”

The passages where he describes the mourning of the prisoners for Woolridge before and after he dies are beautiful. The contrast between the men's shabby surrounds and

the glory of their souls as they keep a vigil on Woolridge's behalf rends us:

“The Warders with their shoes of felt
 Crept by each padlocked door,
 And peeped and saw, with eyes of awe,
 Gray figures on the floor,
 And wondered why men knelt to pray
 Who never prayed before.
 All through the night we knelt and prayed,
 Mad mourners of a corse!”

5. Dual beings

It is possible to believe in both bits of reasoning. People are too beautiful and important to be brutalized, and too fallen to administer punishment without being hypocrites. They're not logically inconsistent, and I don't think they're aesthetically or emotionally inconsistent either. Just like a sufficiently skilled art work can contain moments of appalling ugliness alongside tremendous beauty without those “canceling out”, so too are people woven through with glory and horror. Too beautiful to be judged, and too ugly to judge something as glorious as a human.

6. The ballad and eroticism

I am speculating here, but I wonder if there isn't something erotic in Wilde's outlook on Woolridge:

"And I knew that he was standing up
 In the black dock's dreadful pen,
 And that never would I see his face
 In God's sweet world again.
 Like two doomed ships that pass in storm

We had crossed each other's way:
 But we made no sign, we said no word,
 We had no word to say;
 For we did not meet in the holy night,
 But in the shameful day."

Now, this is further stepping into the realm of pure speculation, but I wonder if that romantically charged perspective on Woolridge wasn't a path by which Wilde humanized him—saw past the horrific thing he'd done? Romantic and erotic energies have this power—to randomly connect us with, and make us sympathizers for, people we would otherwise despise, or at least try not to think about. This is a side of the erotic we don't often consider. We often conceive of eroticism as turning people into objects in our mind, but what about its capacity to make us sympathizers? Sometimes this power takes on a sinister or at least ambivalent aspect—like the people who fantasize about serial killers and court them in prison. Sometimes it is exalted in literature, as in *Romeo and Juliet*: *Would a rose by any other name not—etc. Etc.*

7. Feminism and the ballad

Obviously, a conservative will find much to disagree with in the poem, but “The Ballad of Reading Gaol” is an uncomfortable read, whatever your political orientation. I’m all for mercy, but as someone who thinks women have historically had a rough deal, I wasn’t comfortable with Wilde’s seemingly blithe dismissal of Woolridge’s murder of his wife:

“Yet each man kills the thing he loves
 By each let this be heard,
 Some do it with a bitter look,
 Some with a flattering word,
 The coward does it with a kiss,
 The brave man with a sword!”

“Well sure, he brutally murdered his wife, but in a funny sort of way, doesn’t every man kill his wife?” To which the answer is no. There is a very important sense in which the vast majority of men don’t kill their wives—the literal sense. Is Wilde playing with words here to minimize a gross act of violence against a woman?

Perhaps. But there’s also a sense in which Wilde’s sentiment can be read not as minimization of what Woolridge did, but maximization of the emotional violence inherent in a certain sort of marriage. In this regard, this stanza might be read not as an apologia for Woolridge, but as a biting critique of patriarchal marriage. I’m not fully comfortable with this defense of Wilde, but we shouldn’t feel comfortable about art. To fully draw out the critical power of the poem, we must remember that there are four victims in it. The first is Woolridge; the second is Wilde and the prisoners collectively; the third is the collective warders, doctors, and reverends of the prison who are brutalized by what they do; and the fourth, and most gravely wronged of all, is Laura Ellen/Nell Woolridge, murdered by Thomas Woolridge. Having recognized the victims we need to consider the possibility that merely because they are human, not a single one of them deserved what happened to them.

The passage also has to be read alongside Oscar Wilde’s own life. Wilde was aware of the sour face of love. Love sent him to prison and ruined his health and his reputation.

8. Antipoenatism

Because I like to make up words, let’s call generalized opposition to punishment antipoenatism, from “poena” which is Latin for “punishment” and “anti” which is Latin for “anti”. Could antipoenatism ever be viable? Is antipoenatism the sort of idea which depends for its interest on whether it is, or ever will be, viable?

No. Antipoenatism is pretty obviously not viable with the world the way it is—some people need incentives not to do bad things. However, it could be viable in a future

where we have the technological capacity to restrain the violent without removing their liberty (see Iain Banks' concept of the slap drone) or to cure the violent of their violent tendencies.

But I think antipoenalism is an idea that holds power even in a world where it is not feasible, and should hold that power to shock and shame us all. Jesus' provocation, "Let he who is without sin cast the first stone", holds our imaginations even now. We need a compass that points us towards utopia, even if we can't make it there, and even if it can't be real—you won't see the world as it is without crazy dreams of what it could be.

9. Breaking chains of silken beauty

It's very interesting that the greatest piece of work by Wilde is "The Ballad of Reading Gaol." Wilde was an aesthete, holding that art should be for its own sake—the sake of beauty—and not to serve pedagogic, political or moral purposes. How weird then that his best and most passionate work brims with moral significance and feeling. The chronic ironist driven by circumstances to express real passion is a potent thing (happens all the time on Twitter). I wonder—and this is pure speculation—if Wilde's aesthete sensibilities weren't like a shell to contain his powerful moral sense, which perhaps he feared might be, in today's language, "cringe". When the physical, emotional, and moral torture he had experienced finally burst through that qlippah, his best work emerged.

10. Sorrow for Wilde

I don't know if Wilde was, in any overall sense, a good person, I haven't studied his life closely, and even if I had, I am no judge of souls. But it is unbearable to think of what happened to the spirit, at once both kind and soaring, present in this poem. Fuck you to everyone that valorises the sort of society that did that to Wilde.

A brief reply to Scott Aaronson's- "We Are the God of the Gaps"

This is not a reply in the sense of a refutation- I doubt I will say anything here SA would disagree with. Rather, it is an extension or meditation. Scott Aaronson has written a prose poem called "We Are the God of the Gaps". I quite like it, it occupies an odd space between a traditional prose poem and an essay. The thesis of the poem is that machines will gobble up all human tasks. Finally, the last task remaining, the one thing that a computer will never beat you at, will be predicting exactly what you will do in the very next moment:

"And the rising tide of the learning machines will flood them all, Poker to poetry, physics to programming, painting to plumbing, which first and which last merely a technical puzzle,"

For Scott, this is a melancholy prospect, because it hits at his self-esteem:

"From earliest childhood, those of us born good at math and such told ourselves a lie:

That while the tall, the beautiful, the strong, the socially adept might beat us in the external world of appearances,

Nevertheless, we beat them in the inner sanctum of truth, where it counts.

Turns out that anyplace you can beat or be beaten wasn't the inner sanctum at all, but just another antechamber,"

I get it. I do. Consider progress in chess. At some point, the best player in the world stopped being a human and became a computer. Humans briefly made themselves relevant through the creation of centaur chess, a computer, and a grandmaster working

together to choose moves, often with the computer doing tactics and the human doing the positional play. Then, a few years after the invention of centaur chess, computers became so good that having a human second guess its moves only made it weaker.

Sometime in the next few years- decades at most, something similar will happen in my field of writing thoughtful essays. The best essayist in the world will be a computer. Superb human essayists will briefly hold onto relevance through collaborating with computers, and then they'll be surpassed permanently. Similar things will happen in SA's fields of mathematics and quantum computing.

But I'll live. So long as humanity lives as a whole.

Let me confess to some vanities. I had the immense privilege of being able to make a bid for the status of a great philosopher. Of course, I never had a chance. I've easily met ten philosophers who are better than me. I'd be shocked if I made the top thousand in the world, pleasantly surprised if I made the top 10,000. Still, the world humored me in trying. I got to have my shot, like a very moderately gifted district track and field athlete who dreams of one day being an olympian. It was a joy to tilt at this windmill.

(To be doubly clear- what gave me the chance to take this shot was not any exceptional talent on my part, just a strange autistic narcissism that told me it was possible.)

Scott Aaronson did one better. He not only got to make an attempt at greatness, but he also succeeded. With an H-index of 51, he got to be one of the greatest quantum computer scientists of our age- and since quantum computing is so new, of any age.

Most people never get that. Most people never get to make a real attempt to be recognized as among the very best at something that matters. The number of people who do get to make such an attempt has been shrinking since the stone age. In a band of 10 to 30 people quite possibly most people got to be the best at something that

mattered- singing, fire-starting, painting, good looks, tool making, shelter building, foraging, hunting, fighting, charm. Then as the world got bigger...

Scott's melancholy, a melancholy I share, is an example of what I have previously called writerly bias. The tendency of people- especially readers, but to a degree humanity is a whole, to perceive the world disproportionately through the mentality of a writer, because the people who do the writing are all writers. They all have a writer's problems, solutions, desires, background, etc. A lot of writers get to make an (almost always quixotic) attempt at greatness. Far more than the general population.

Thus a lot of humans are feeling the way Scott describes about AI right now, but that's because this emotional complex is being pushed on them by writers. The bigger issue- rarely talked about- is what most of humanity has already lost- not just the status of being "the best" at anything, but even the chance to make an attempt at being "the best" at anything.

The vast majority of people are already living in a life-world in which attempting to be the best at something is unthinkable. That's something that was taken from them perhaps sometime between 8000 BC and 1 AD.

Scott and I's shared melancholy is a melancholy not of the beginning of a process of human loss, but of the moment of its final consummation. The end of a march that began in deep time. A march towards fewer and fewer people, as a proportion, being the best, or among the best, or even making an attempt at it.

Everywhere you go, you always take the weather with you

“Everywhere you go you always take the weather with you”

-Crowded House “Weather With You”.

In this essay I want to talk about philosophy and OCD. Not for the purpose of creating a “philosophy of OCD”, but rather to find ways in which the vantage point of OCD can aid philosophical inquiry.

An obsessive and compulsive life

Obsessive compulsive disorder is a horrible thing and I would not wish it on anyone. That said, living with OCD has given me a certain way of seeing the world. The insights and viewpoints of that way of seeing are not unique to OCD, but OCD is certainly one path to them. I decided to write this to share what the refining fire of a lifetime with OCD has given me.

1. Generalities

Obsessive compulsive disorder is defined by the presence of one of, or both of, obsessions and compulsions. It is not necessary to have both obsessions and compulsions in order to have OCD– for example, some sufferers either have no compulsions, or very subtle compulsions that only become obvious once the context of the obsessions is understood. By the DSM V an obsession is defined as follows:

1. Recurrent and persistent thoughts, urges, or impulses that are experienced, at some time during the disturbance, as intrusive and unwanted, and that in most individuals cause marked anxiety or distress.

2. The individual attempts to ignore or suppress such thoughts, urges, or images, or to neutralize them with some other thought or action (i.e., by performing a compulsion).

And a compulsion is defined as:

1. Repetitive behaviors (e.g., hand washing, ordering, checking) or mental acts (e.g., praying, counting, repeating words silently) that the individual feels driven to perform in response to an obsession or according to rules that must be applied rigidly.
2. The behaviors or mental acts are aimed at preventing or reducing anxiety or distress, or preventing some dreaded event or situation; however, these behaviors or mental acts are not connected in a realistic way with what they are designed to neutralize or prevent, or are clearly excessive.

There are a few other parts to the definition, but we can safely leave them aside for our purposes.

The obsessions and compulsions can take endless forms. The public is most familiar with fears related to germs and hygiene, but there are numerous other types, ranging from fear of offending God, to fear of being a pedophile. My own type, which we will discuss in the next section, is the fear of harming others.

3. My specifics

“Anyone who doesn’t have a guilty conscience needs one, and anyone who does have one, doesn’t need one.”

–Something I once read on a Magic the Gathering Card

One more bit of general background—the kind of OCD I suffer from. I suffer from a form of OCD that makes me afraid of enacting physical, sexual or psychological violence on others, sometimes called “Harm OCD”. I fear both the harm I might do to others, and the consequences I might face as a result of doing such harm. Although this form of OCD is quite widely understood among experts, not many people in the general public are aware of it.

Partly I think this is because of the nature of the content of this form of OCD—it’s self-censoring. Although it’s embarrassing, you might be willing to admit that you’re afraid of catching aids in a public bathroom for example. However, admitting that you’re afraid of sexually assaulting someone, or kicking a child in the face as you walk down the street, is scarier. Thus knowledge of this form of OCD hasn’t seeped into the public consciousness.

A lot, but not all of my fears, centre on the following nexus. Within this nexus there are numerous variations on a theme and elaborations, and I am perhaps misrepresenting what is going on in some respects, but this is the basic story. Suppose a random thought of groping someone on the subway pops into your head. At first, so long as you recognise that they are not really your thoughts or desires this thought may not be very frightening. Suppose though that you began to worry that because such thoughts had popped into your head so many times, that you might perform these actions on autopilot—mechanical reflex—without even being aware of it, let alone intending it. You now have a plausible story about why you should be afraid of these thoughts, and also a story about why being afraid of these thoughts is dangerous in itself. Trying to be rid of these thoughts becomes like trying not to think of a pink elephant. You will think these thoughts all the more because you perceive them as dangerous, and in your mind that makes them even more dangerous.

Here is a simplified schemata of the process:

Thought of harming someone>>>Thought that you've had that thought so many times you might do it on autopilot>>>The thought is now perceived as dangerous>>>This makes you want to stop thinking about it>>>This makes you think about doing it more>>>Loop continues

Because you're worried about doing these things 'automatically' you don't even have the comfort of reasoning 'well nothing has happened yet and it's been a long time, so it should be okay'. Part of what you're afraid of is that you might be doing such things all the time, and just not noticing. Perhaps the victims were perhaps too shocked or scared to say anything.

The irony of harm OCD is that all available evidence suggests that people with it are unusually appalled by violence, and thus much less likely to commit it than the general population. There are no recorded cases of someone with OCD committing violence unintentionally, or in a way in which relates to their fears.

4. The Red Queen Hypothesis of OCD: Fear as genetic algorithm

"Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"

-The Red Queen, Through the Looking Glass

One of the things I find interesting about OCD is the process by which it generates fearful obsessions perfectly targeted against its subjects- expertly finding their weaknesses. In coming to understand this, I realised that OCD is a self-generated mental parasite, operating like an evolutionary algorithm.

Fearful thoughts are generated (somewhat) randomly, those fears that are successful (in the sense of capturing attention and thus cognitive resources to elaborate on them) are selected—just as with a genetic algorithm. These fears give rise to lineages of similar fears and variations on a theme which are in turn elaborated on. Over time, a thoughtful

sufferer will come to see holes in their fears, and this leads to arms race where the fears evolve in response to keep up with the growing dialogue between fears and defences, like the red queen theory of evolution in biology, or, like the runaway evolutionary processes of cancer, constantly subverting and being subverted by our natural defences.

Coming to understand that a part of me—my fearful thoughts—was evolving and acting with logic that was blind to, and actually opposed to my own purposes was terrifying, like finding an eldritch horror in your own breast. However, it was also a breakthrough not only in my struggle with OCD, but in my struggle to know myself.

5. You don't really know scepticism until you've fought delusions

You usually learn facts through indirect sources, and you know that these sources can be faulty, even if you consider that unlikely. On the other hand the basis of your reasoning about, and inferences from, those facts is immediately present to you. If you don't find an inference convincing, you will most likely drop it. Thus we have a cognitive blindspot when it comes to being sceptical about our own reasoning—doubting the premises on which we build our inferences is easy, but doubting your own reasoning—the inferences from those premises—is hard, because it wouldn't be your reasoning if you weren't already directly convinced it was persuasive. Thus there is an asymmetry in our capacity to doubt our premises and our inferential moves- the former is easier than the latter.

Your accepting a sequence of deductive inferences as valid is partly based on your inability to imagine ways it might be wrong. Even in the case of inductive reasoning, your accepting a sequence of inferences as valid is at least based on it being hard to imagine a plausible scenario in which the premises hold but the conclusions don't. On the other hand, imagining a factual premise is wrong is easy—"I misremembered" or "Maybe someone gave me wrong information". This asymmetry is a real shame, because as many errors are made through faulty inferences as through faulty premises, so we should be just as sceptical of our reasoning as our premises.

A few concrete examples. Witness the incredulity, confusion and even anger which can arise when someone is corrected about the Monty Hall problem. Another example is the way in which many scientists worry about whether their methodology is generating reliable data, but then often scarcely worry at all about whether there are any gaps in their method of statistical analysis.

I believe OCD can help with this—and maybe even if you don't have OCD reflecting on the experience of an otherwise reasonable person with OCD can shake false confidence in your inferential capacities. Having been in a state where you feel like everything is about to come crashing down on you again and again, only to find each time that your carefully manufactured chains of inferences about what would happen next meant nothing even though you could see nothing wrong with them is very educational. In my experience OCD typically takes a series of commonplace facts and real observations and weaves a story which seems perfectly plausible to the sufferer according to which the only reasonable inference from these facts is upcoming disaster. The inferences involved seem very persuasive, but turn out to be meaningless—so you have to become good at recognising that there is no link between how persuasive it seems to infer D from A, B, & C, and the real validity of that inference.

Recognising that an individual chain of reasoning can always be invalid—no matter how good it looks—is the first step to becoming a hedgehog. More on that later.

6. Epistemic judo—turning doubt in upon itself

“Doubt your doubts...”

-Switchfoot

I began to articulate the idea of epistemic judo. In real Judo, you leverage your opponent's weight against them. In epistemic judo, the person afflicted with doubt and inconfidence leverages those feelings against their own anxieties. Empirical evidence

suggests that people with OCD are unusually sceptical and doubtful, even about things not related to their fears. Doubt and the trait of being doubtful seems to be intimately related to the aetiology and processes of OCD, leading some to call OCD the disease of doubt. How could the OCD sufferer leverage this capacity to doubt everything against their own fears?

The first step is to think of your mind as a mechanism for generating understanding of the world. Return to our earlier model of OCD as an evolutionary algorithm for finding blindspots in your epistemic processes, and using them to trigger fear. You can try to beat it by seeing why every single lie it tells you isn't plausible after all, and this will help to some extent, but to make progress past a certain point you have to stop believing that just because a fear seems plausible, it is plausible—even if you haven't got a great counter argument for it yet. If you can't take this step, your OCD will keep throwing up new fears and variations on the old fears faster than you can spot what is wrong with them. In other words, you have to stop believing that your own OCD driven cognitions are reliable, even if you can't see the specifics of why they are wrong. You may not have found the logical gap in the argument that you should be afraid of X yet, but eventually you will, and even if you don't, chances are that it exists.

The strategy to achieve this is to turn the sceptical/doubting mindset OCD generates against itself, until you reflexively doubt your fear arguments. To a certain extent, you have to stop believing in your own ability to reason about topics that overlap with your OCD. Ordinary people have trouble engaging in blithe and automatic scepticism of plausible sounding ideas they themselves have generated. However, people with OCD have so much practice questioning and doubting everything, what is it to doubt one more thing? Eventually I turned “I am not qualified to think about these topics well or impartially” into a mantra.

7. Arguing with your feelings

“Thoughts aren't facts”

-Common Parable

“Feelings aren’t facts”

-Another common parable

Most narratives and most bits of folk wisdom, tell us to trust our feelings—our “intuition”. Perhaps the hardest part of grappling with OCD, and one of the most educational, is learning not to trust your intuition. Sometimes feelings are right and sometimes they are wrong. Just because you feel like SOMETHING IS VERY WRONG doesn’t necessarily mean shit. If you want to overcome OCD, you have to learn to be skeptical of- and sometimes even outright ignore- your own intuition. If you think of feelings as unlike thoughts—as beyond debate, let alone volition—you will find this difficult. Hence you must come to understand the similarities between thoughts and feelings. Both are rooted in ideas about how the world is, and neither are immune from criticism. For example, being “afraid” of a spider is almost always accompanied by the thought that it can hurt you, or crawl all over you in an unpleasant way. You feel the way you do because you have a certain model of the world. There is no sharp distinction between feelings and thoughts.

The wrong way to think about this is to think of it in terms of feelings and emotions being always inferior to formal reasoning—of feelings as inferior or flawed thoughts. Actually, some feelings are very accurate. Sometimes the explicit chains of reasoning we call ‘thinking’ are much worse than the pattern matching we think of as feelings or intuitions, and sometimes the opposite is true. This is unsurprising because, again, emotion was never discontinuous with reason, and both feeling and reasoning are subject to critique, revision and scrutiny.

People have accused me of valorising the Cartesian thinking subject at the expense of the embodied subject of feminist epistemologies yada yada with this view, but I think the opposite is true. Only by challenging the false separation of thought and feeling can we see what is wrong with both the mystical valorisation of intuition and the obnoxious

assertion that one has transcended reason for pure rational assessment. “Trust your feelings” and “Suppress your feelings so you can be truly rational” are two sides of the coin, for they both posit a separation, and disagree only on which has priority.

8. Becoming a hedgehog: OCD and the struggle against single model thinking

“Homo unius libri timeo”

-St Thomas Aquinas

Almost all OCD type fears have the following structure—a prediction about how the future will be (or in the case of guilt, how the past was) based on a specific sequence of events in a causal pattern. At least one of A-Z must happen and then this will follow this and then either this or that will happen but then in either case necessarily this must precede... and so on. Even though every individual step might sound plausible, something almost always breaks down, and so the fears of people with OCD are almost never true. This form of unitary model thinking is an extreme type of the “Hedgehog” cognitive style identified by Isaiah Berlin, and contrasted with the “Fox” cognitive style, based on a more flexible appraisal of different factors and plausibilities. Another name for these modes of thought is Euclidean versus Babylonian methodology.

There’s a great deal of evidence that the hedgehog type approach breaks down not just in OCD, but in things like expert attempts to predict future events. Overall being a hedgehog isn’t a great strategy (although beware anyone who tells you it is always wrong—we wouldn’t want to be meta-hedgehogs).

As I saw how being a hedgehog was making me mentally ill, I made an effort to become more of a fox, not just with regards to OCD, but throughout all of my life and activities. For example I tried to view the challenge of predicting the future not as a matter of charting a sequence of events, like falling dominoes or a Rube-Goldberg machine, but

instead tallying power factors, tendencies, resources, general drifts of various coalitions and past trends, and throwing in a generous dollop of pattern matching.

9. I am not special: Taking the outside view on yourself

There's a kind of soft narcissism that OCD forces on you. It doesn't help that OCD makes you live inside your own head and spend so much time talking to yourself. Your problems seem very distinct from the problems of other people who have your conditions. Their fears sound so absurd, whereas your own sound so reasonable to you (due to having been specially crafted for your cognitive blindspots). It can seem like the standard treatments could never help—e.g.: dammit I don't need to relax, I need to establish that I won't grope someone! Or: What's the point in going to a psychologist—I'm just as clever as they are, any point they can make about how my fears are illogical I will have certainly thought of myself! Yet all of this turns out to be false, chances are you're really not so different from other people, and will benefit from, and be harmed by, roughly the same things as them. Certainly that was my experience. Recognising that you have all the fallibility—and strength—of other people is incredibly liberating. It's okay to reason as follows:

“Hey, this stuff seems plausible to me, but I am a delusional person, so rather than using what the base odds seem to me to be, I should think about it from an outside point of view. How many people with OCD feel their OCD fears are plausible (almost all), and how many actually turn out to be right (almost none, and literally none in cases of harm OCD like you have). Therefore it's massively unlikely that you will be the first person with harm OCD ever to have your fears realised. The fact that you think you are an exception to this rule, doesn't matter a damn, because there is a 100% chance you would think this, given your OCD.”

In other words, it's okay to take the outside view on yourself. For example, assessing the plausibility of your own thoughts just like a clinician who was only vaguely aware of the content of your thoughts, but knew your diagnosis, would assess them.

This isn't just about the plausibility of your fear arguments. You have to take the outside view on a lot of lifestyle and treatment factors. For example, "While it seems that my problem is the dangerous situation I'm in, so sleep won't help, if I actually get some rest, the experience of countless humans over many millennia indicates the situation will probably seem very different". Or "While I don't feel consciously lonely, and socialising doesn't feel like a priority since I believe my life is about to collapse, the experience of the human race as a whole would suggest it's a bad thing that I haven't talked to anyone except my parents in almost three days and that this may be the real reason for my terror. I need to go meet up with a friend".

I've taken this attitude and applied it to other areas of my life. I don't regard my studies and writings as attempts to find truth in any personal capacity because the odds that I will discover something important are much lower than the odds that my novel ideas are just crankery. Instead my hope is that I will contribute to the social process of truth seeking. I look upon my own work indulgently—from the outside perspective it is true that my work is likely air and puffery, but also from the outside perspective it's true that you could say that of just about anyone so it is important that we don't let the fact that our work will almost certainly be either wrong or unoriginal stop us.

10. OCD and non-self

"I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement."

-David Hume

It's kind of pat, but it's got to be said. At a certain point, gazing at the changes within yourself, and understanding them as intertwining and sometimes contending forces you begin to see the inside of your own brain not so much as a monarchy with an ego-king,

but as a republic, full of traitors, excitable mobs and hard working but beleaguered bureaucrats. Seeing yourself as a balance of forces can inspire vertigo, but it can also inspire hope. A unitary subject with attributes attached is hard to change, a contending swarm can shift.

Perhaps the best metaphor is the self as an ecology. Particular thoughts and feelings you have are like interacting species lineages, each fighting for survival. Drives are like the terrain. Perceptions are like the weather, carrying in nutrients for the organisms and taking them away. You are this jungle- this tangle of evolutionary game theory.

Less abstractly, I remember sitting in a lecture when I was much younger, shaking with terror that I could scarcely conceal. I began to imagine myself as like a boat on a stormy sea, except because I was both the boat and the sea I could never escape. Wherever I went I would always take the weather with me. As painful as this moment was, it stayed with me because it was the beginning of insight.

Later, I beheld the mirror and saw nothing reflected there, and so realised that “I” wasn’t beholding the mirror at all, I was the mirror, I was the room in front of it, and I was the wall it stood on. There is no person in that room, and there could be no person in it, because that room is a series of components which make a person.

Granting textual authority to overcome textual tyranny

The point that many religious and political movements take their texts seriously rather than literally has been made by numerous people before. Here, I want to insist on a small addendum, to this point: taking a text less literally can sometimes be accomplished by taking it more seriously. I talk about three different movements- Rabbinic Judaism, academic Marxism, Catholic Christianity and American civic religion and argue that each of them has used a different strategy, but with the same outcome- it's not just that they take the text seriously but not literally, it's that the very strategy which lets them take it more seriously also enables them to take it less literally.

That is, I'm going to talk about how religious movements, and similar groups, overcome their holiest texts through veneration. By treating them as supramundane texts, they make the question of what the texts merely appear to say on an ordinary way of reading moot. We'll find however that the diversity of ways through which they do this matters as much as their unity.

Perhaps a catchier way to put this is to say that they overcome the tyranny of the texts by increasing the authority of the texts. We've mostly forgotten about this one weird trick in 2021, because we view taking things seriously as the same as taking things literally. The idea that we can take something seriously without taking it literally might be just imaginable, but the idea that you can go further and avoid taking it literally exactly by a strategy of taking it seriously? That idea is outside how we normally think of things.

I. A lost debate, an aside

When I was in primary school I was briefly in the primary school debating team, (a habit I, fortunately, did not continue in high school). The proposition was given to us- and we

had weeks to prepare- That the land of Australia should be returned to its indigenous peoples. We were the affirmative.

We thought this question was impossible for the affirmative. Less than 5% of Australians are indigenous. Handing over governance to such a small minority would be impractical and immoral. The injustices of the past were great, but the egg could not be unscrambled.

Through a very dubious reading of the dictionary entry for the word “returned” we developed roughly the following argument. Australia should be returned to its indigenous peoples in a *spiritual sense*. We thought this subtle distinction would win us the debate- our opponents would be entirely unprepared.

Our opponents simply ignored our creative redefinition of the question and proceeded to demolish the idea of literally handing over a nation of 20 million people to the control of a minority well less than a million. Meanwhile, we spoke past them completely and spoke glowingly of the need for “spiritual” sovereignty.

We lost the debate deservedly. It was a shit question to give an 11-year-old though. Since then, I have often wondered if there is, in at least some contexts, a way of respecting things “in a spiritual sense” whilst avoiding literal meaning that is not, at heart, infidelity. I still wonder about this. This essay won’t answer this question, but the question does frame it.

II. Interpretation in the context of a tradition, an introduction using Rabbinic Judaism as an example

To reiterate, this is not a general essay on the concept of interpreting a text in a tradition. It is an essay that argues specifically that one strategy to get around difficult texts is to interpret them as so sacred that, paradoxically, what they appear to say

matters very little- at least by the standards of a modern reader. However, for those readers without a general background in how texts are interpreted in religious and political traditions, it might be useful to sketch out a broad schematic of how traditions founded on a text work.

The idealized, pure form of what I am talking about is probably best seen in generations of Jewish interpreters. Rabbinic interpretation is an alliance between the living and the dead. The living want the flexibility to interpret tradition in a way that serves the needs of their community, their view of the good, and their particular theological or political hobbyhorses. The dead want the authority of what they have written to be upheld, and its wisdom to be recognized. The dead possess the authority of generations. The living has a power the dead do not have, they can speak and respond directly to the existing situation.

Thus the living and the dead make a pact. The living will respect the authority of the dead, the source texts, the commentaries, the commentaries on the commentaries. The dead will allow their words (like they have any choice) to be reappropriated to support the agenda of the living.

Orthodox Jews believe in what is called *Yeridat ha-dorot*, the decline of the generations. Each generation has less authority than the last. On the surface this might seem terribly cruel to the living- are there not wise and good people among them? Yet on reflection, elevating the dead in such a way may be necessary to give them any power at all, for the living have a great power indeed the power of getting the last word in. Or at least the last word for now. Thus to let the dead participate in the conversation at all on anything even like an even footing *yeridat ha-dorot* may be necessary.

The dead are powerful in authority but rigid, the living are weak in authority but supple. Their alliance is powerful. And it goes on in a cascade. Those now living will soon be dead and the alliance is made again in each generation. Perhaps the knowledge that

they will soon be dead motivates the living to treat the dead well, for any callous precedent they establish of ignoring the dead will be applied to them in turn.

As alliances go, it is rather like one from the world of international diplomacy. There are public statements made of mutual respect, but underneath this, there is tension, hopefully constructive. The word “Israel” means “Wrestles with God”. While on the surface, all is smiles, the past is not just something to be let into the present in the way a child might, through a simple interpretation of the text. If, as is often said, the past is another country, strict border controls between past and present are necessary. The text must be wrestled with, slavish application would be nearly as great an error as discarding it altogether.

The text that usually mentioned when explaining the nature of religious authority in Judaism is found in Bava Metzia 59a-b- the famous Oven of Akhnai:

“After failing to convince the Rabbis logically, Rabbi Eliezer said to them: If the halakha is in accordance with my opinion, this carob tree will prove it. The carob tree was uprooted from its place one hundred cubits, and some say four hundred cubits. The Rabbis said to him: One does not cite halakhic proof from the carob tree. Rabbi Eliezer then said to them: If the halakha is in accordance with my opinion, the stream will prove it. The water in the stream turned backward and began flowing in the opposite direction. They said to him: One does not cite halakhic proof from a stream.

Rabbi Eliezer then said to them: If the halakha is in accordance with my opinion, the walls of the study hall will prove it. The walls of the study hall leaned inward and began to fall. Rabbi Yehoshua scolded the walls and said to them: If Torah scholars are contending with each other in matters of halakha, what is the nature of your involvement in this dispute? The Gemara relates: The walls did not fall because of the deference due Rabbi Yehoshua, but they did not straighten because of the deference due Rabbi Eliezer, and they still remain leaning.

Rabbi Eliezer then said to them: If the halakha is in accordance with my opinion, Heaven will prove it. A Divine Voice emerged from Heaven and said: Why are you differing with Rabbi Eliezer, as the halakha is in accordance with his opinion in every place that he expresses an opinion?

Rabbi Yehoshua stood on his feet and said: It is written: “It is not in heaven” (Deuteronomy 30:12). The Gemara asks: What is the relevance of the phrase “It is not in heaven” in this context? Rabbi Yirmeya says: Since the Torah was already given at Mount Sinai, we do not regard a Divine Voice, as You already wrote at Mount Sinai, in the Torah: “After a majority to incline” (Exodus 23:2). Since the majority of Rabbis disagreed with Rabbi Eliezer’s opinion, the halakha is not ruled in accordance with his opinion. The Gemara relates: Years after, Rabbi Natan encountered Elijah the prophet and said to him: What did the Holy One, Blessed be He, do at that time, when Rabbi Yehoshua issued his declaration? Elijah said to him: The Holy One, Blessed be He, smiled and said: My children have triumphed over Me; My children have triumphed over Me.”

But note -and perhaps I am overstepping my bounds here- that if the law is not in heaven, it is certainly not in the past either. It is here. Obviously, tradition arrives to us from the past but right now, the decisions about how it applies are being made here, on the basis of what people here say. The authority of the tradition must be upheld, the flexibility of present must be upheld, and so it is necessary to wrestle with the law, and with one’s own temptation to rule too leniently out of indulgence, and one’s own temptation to rule too strictly, out of caution.

I have used Judaism as an example, but although none fits the type quite so well, other interpretative traditions in a broadly similar mould exist.

III. The origins, and inversion, of politics

Many, though not all, hunter-gatherer societies are very egalitarian. An anthropologist named Boehm suggests we understand this in terms of counter dominance hierarchies. Here is my gloss on this idea:

These early hunter-gatherer societies weren't simply passively egalitarian, rather their egalitarianism was an achievement that had to be constantly upheld. The reason that human bands are often egalitarian, whereas great ape bands often aren't, is that humans developed politics. In its original form, politics was a weapon of counter-power, not power. By allowing coordination by groups of weaker individuals to overcome stronger individuals, politics created early human egalitarianism. Only later was the original use of politics inverted, then it became a tool for an individual to control much larger groups than any alpha chimp or silverback gorilla could have ever dreamed of.

However, per Boehm, politics in hunter-gatherer societies never overcame hierarchy "once and for all". The tendencies towards hierarchy were always there. Rather, power and counter-power became locked in a struggle of millennia. Powerful individuals tried to consolidate their position, but the group, both consciously and unconsciously, pushed back. The result is that hunter-gatherer societies became riddled with institutions designed to prevent the emergence of powerful individuals and minority factions. These institutions ranged from mocking successful hunters and downplaying their achievements to discrete assassinations of fellows who wouldn't take the hint.

I am reading David Graeber and David Wengrow's book "The Dawn of Everything" and that has me fascinated by the cunning and strategy with which ancient societies were able to play this game of power and counter-power. It's like a game of four-dimensional chess, by correspondence, with time per move allowed on the scale of generations. The struggle to restrain tyranny- tyranny of any kind- is like a subtle and invisible war that pervades history. A common tactic is the decoupling of power and authority. We see

power without authority- for example- police officers who were also ceremonial clowns. We also see authority without power- esteemed hero chiefs who are forbidden from actually ordering anyone to do anything and are required to be the poorest people in their village through constantly providing for everyone.

IV. Undermining power with authority

Today we are used to the idea of constitutional monarchs who are not really allowed to do anything. Zoom out a bit though, and it's quite weird. The idea of a person who, theoretically, is in charge of everything, but in practice, dare not do anything at all, even in some cases appoint their own servants, seems like it should be a very unstable arrangement- like a constitutional crisis waiting to happen. Even should it prove to be stable, it seems like a waste of resources.

Yet historically it is quite common- it is certainly not just an innovation of the modern period. We find long-lived instantiations of this configuration in dark ages France, and in medieval Japan, for example. Moreover, there are often tendencies in this direction, even when they are not fully realized. Many kings and queens throughout history gave up enormous amounts of their power to advisors.

I'm not a historian of monarchy, but I want to hazard a guess as to why this social form might be relatively stable. It resolves the tension between kingly power and kingly authority. In a manner of speaking, authority can be seen as being right in the abstract. Power can be seen as making the decision about what to do. The problem is that these aren't very compatible in the long run for mortals. When we make a lot of decisions, it quickly becomes obvious that we are not always right. So to preserve the appearance of kingly authority, kingly power is stripped. Thus in some, though admittedly not all, circumstances, precisely as kingly authority reaches a zenith, it becomes necessary for kingly action to come less and less.

But what does it look like to put a text in a similar gilded cage?

V. Example: Rabbinic Judaism

The method by which Rabbinic Judaism overcomes literalism is through superabundant meaning. The text is laden with so many truths, layers of meaning, and specificities that what appears to be “at first glance” on a literal interpretation may, in fact, not be true or be only partially true. The text is made more authoritative and profound- by emphasizing that nothing in it is accidental and it contains vast meaning- even at the same time as a casual literalism is discarded.

Rabbinic Judaism has a deep respect for the Torah. It is said to have been written before the foundations of the world, according to one source 974 generations before its foundations. It is said to have served as the plan for the design of the world. Despite this (or rather, as we will see in part because of this), rabbinic Judaism is famed and sometimes criticized, for the extremely creative ways it interprets the Torah.

Often the changes are in the direction of humanitarianism. A great example of how the Rabbinic tradition “softens” what the source material apparently says is in the laws concerning the death penalty. The Rabbis end up coming to the conclusion that, in practice, almost no one should ever be put to death under God’s law for almost any reason whatsoever.

And I do not mean “one in a thousand” or even “one in a hundred thousand” I mean almost no one. One Rabbi, R Eleazer even remarked that it is a bloody Sanhedrin that kills a man even once in seventy years (i.e. at all, ever). I won’t go through all the processes by which they reach this conclusion, for they are many, but let me share one example.

In Leviticus 20:9 it is written: “Anyone who curses their father or mother is to be put to death. Because they have cursed their father or mother, their blood will be on their own head.

In Deuteronomy 21:18 -21 we get a bit more detail:

“If someone has a stubborn and rebellious son who does not obey his father and mother and will not listen to them when they discipline him, his father and mother shall take hold of him and bring him to the elders at the gate of his town. They shall say to the elders, “This son of ours is stubborn and rebellious. He will not obey us. He is a glutton and a drunkard.” Then all the men of his town are to stone him to death. You must purge the evil from among you. All Israel will hear of it and be afraid.”

The Christian interpretation of this has generally been that a nation under Mosaic law should put rebellious kids to death. Some of them think this reflects a kind of “ideal law” of the old covenant, and is not something that was ever really meant to be put into practice- this would probably be the majority view among theologians. In any case, they hurriedly add, we are under the new covenant, not the old, so it no longer applies. At the other end of the spectrum, some Christian dominionists for example, think that we should literally stone to death a bunch of teenagers. Now, to be fair, many caveats will be insisted upon “the text is talking about really rebellious kids here, we’re not talking about a casual slip of the tongue” etc., etc., but ultimately they do think that at least some teens should be killed for rebellion.

The Jewish approach is quite different. A close and imaginative reading of the text makes the offense so specific that it would never actually apply to anyone. The Encyclopedia Judaica has it:

“Interpreting every single word of the biblical text restrictively, the talmudic jurists reduced the practicability of this law to nil. The "son" must be old enough to bear criminal responsibility, that is 13 years of age (see *Penal Law), but must still be a "son" and not a man: as soon as a beard grows ("by which is meant the pubic hair, not that of the face, for the sages spoke euphemistically") he is no longer a "son" but a man (Sanh. 8:1). The period during which he may thus be indicted as a "son" is three months only

(Sanh. 69a; Yad, Mamrim 7:6), or, according to another version, not more than six months (TJ, Sanh. 8:1). The term "son" excludes a daughter (Sanh. 8:1; Sif. Deut. 218), though daughters are no less apt to be rebellious (Sanh. 69b–70a).

The offense is composed of two distinct elements: repeated (Sif. loc. cit.) disloyalty and defiance, consisting in repudiating and reviling the parents (Ex. 21:17), and being a "glutton and drunkard." This second element was held to involve the gluttonous eating of meat and drinking of wine (in which sense the same words occur in Prov. 23:20–21), not on a legitimate occasion (Sanh. 8:2), but in the company of loafers and criminals (Sanh. 70b; Yad, Mamrim 7:2) and in a ravenous manner (Yad, Mamrim 7:1). There are detailed provisions about the minimum quantities that must be devoured to qualify for the use of the term (cf. Yad, Mamrim 7:2–3). As no "son" can afford such extravagance, the law requires that he must have stolen money from his father and misappropriated it to buy drinks and food (Sanh. 8:3, 71a; Yad, Mamrim 7:2). "Who does not heed his father and mother" was interpreted as excluding one who does not heed God: thus, eating pork or other prohibited food, being an offense against God, would not qualify as gluttony in defiance of parents (ibid.). But it was also said that one who in his use of the stolen money performed a precept and thus heeded his Father in heaven could not be indicted (TJ, Sanh. 8:2).

As father and mother have to be "defied," to "take hold of him," to "say" to the elders, and to show them "this" is our son, neither of them may be deaf, dumb, blind, lame, or crippled, or else the son cannot be indicted as rebellious (Sanh. 8:4; Sif. Deut. 219). Either of them could condone the offense and withdraw the complaint at any time before conviction (Sif. Deut. 218; Sanh. 88b; TJ, Sanh. 8:6; Yad, Mamrim 7:8). [EDIT BY PHILOSOPHY BEAR: From memory, some commenters say that because the father and mother both have to say it, they must speak exactly the words mentioned in the verses perfectly in unison].

The son had first to be brought before a court of three judges (see *Bet Din) where, when he was convicted, he would be flogged and warned that unless he desisted from his wanton conduct he would be indicted as a rebellious son and liable to be stoned; if he did not desist, he would be brought before a court of 23, including the three judges who had warned him (Sanh. 8:4; 71b; Mid. Tan. to 21:18; Yad, Mamrim 7:7). If he escaped before sentence was passed, and in the meantime his hair had grown, he had to be discharged; but if he escaped after sentence, he would be executed if caught (Sanh. 71b; Yad, Mamrim 7:9).”

It is precisely by interpreting every single aspect of the text as of vital importance that they, with respect, in a certain, particular sense, render it meaningless. The exact opposite of how a modern liberal theologian would proceed. Am I saying today’s liberal theologians should go back to this older way of doing things? No, it’s beyond me to have an opinion that, but I do think the alternate ways of doing things are well worth thinking through in detail.

Notice how the analysis goes. It’s difficult to follow translated into English, but a lot of this is coming from placing an overwhelming degree of meaningfulness on every aspect of the text- gender and age indicators, adjectives, quantifiers, etc. etc. What might seem to be the plain meaning of the text is undermined by taking every single aspect of it as entirely serious and non-accidental.

The Rabbis were fully aware that so specified the law was very unlikely to be applicable to anyone in the whole history of the world. The conclusion they reached was that this law was included as a warning about how serious problems can get if they are left to fester. Thus the text is transformed into allegory, but not in the usual way that is done in our era- by treating the text as below the level of literal meaning, rather the text is elevated by ascending through higher and higher degrees of literalness.

It's not just about the law either. A literal reading of the story of King David, taking its "plain meaning" would suggest that he:

Betrayed the nation of Israel and fought for its enemies. Indeed, with eagerness.

Committed adultery (and possibly rape as there is no indication given that Bethsheba returned his affections)

Stole a man's wife

Killed that man

In other words, he comes across as a wicked brute who was, for unclear reasons, nonetheless favored by God.

The Rabbinic tradition, through a combination of very inventive and close readings of the text and supplementations of the available material, holds that Davids's sins in this and other matters were extremely minor. He did not actually commit the sins that he seems to commit in the story. Indeed the Rabbis state "whoever said David sinned is surely in error (Babylonian Talmud Shabbat 56a).

This process of interpretation may seem sad to our eyes if we are enraptured of the picture of David as a broken sinner who nonetheless received God's grace. However, there is actually something humane about interpreting David not as a brutish thug. For if David is a brutish thug who is nonetheless beloved by God, it follows that we have no right to judge a king for his actions, because he, like David, might be beloved anyway. By insisting that David was good, they implicitly bound future kings to be like David.

A friend told me once that there's a Christian congressional bible study that tells the representatives and senators to think of themselves as like King David. Just like he did,

they sin grievously, but ultimately God loves them because God chose them for his purposes. Just like king David, their personal indiscretions, regrettable as they may be, are beside the point. Perhaps this brings home how terrifying it could be to have a senator who thinks God wouldn't be all that fussed if they killed a man to marry his wife- they could still be God's favorite.

I have no idea whether any of that about the congressional bible study is true, but I can certainly imagine a wicked king comforting himself with the thought that he was like David. Perhaps then the Rabbis were wise indeed to insist the story is not what it seems. At any rate, a reduction in literalism doesn't just coexist with an increase in authority- rather the increase in authority necessitates the reduction in literalism.

VI. Example: Marxism

The method by which a certain strain of Marxism overcomes literalism via veneration is through conceptual abstraction. The point of the text isn't the concrete claims, it's the brilliant subtleties in the relations between ideas, the way of their unfolding, and also the method by which they are arrived at. The book is portrayed as more brilliant -capturing layer after layer of human relations- at the same time as a surface literalism is denigrated.

Within the Marxist tradition there are many approaches to the classical texts. I think probably the three pure "types" are:

The 1930's Marxist sect approach: A mechanical Marxist literalism about everything Marx (and Engels, and Lenin) say, treating it as a series of more or less quantitative laws etc.

The sane approach: An attitude that treats Marx as a useful thinker who's sometimes right and sometimes wrong, and denies the validity of a straight-up "appeal to Marx" to prove anything

The academic approach: A quasi Hegelian Marxist mysticism that takes him as saying something very profound, but not in the first instance, mundane statements about anything so plebian as actual specific quantifiable things. Rather he's engaged in a profound kind of dialectic, to which what appear to be the specifics he's talking about are dispensable.

I would say that my own approach to Marx is 60% the sane approach, 25% the academic approach and 15% the Marxist sect approach. Mostly I think of him as just the best among the social theorists. Sometimes though I get the sense that he has a special way of seeing and thus I buy a little bit into the academic approach. Sometimes, often in the context of polemics, I find it useful to frame things in terms of the mechanical approach. The mechanical approach is ultimately an oversimplification, but it can be useful to shock someone into another way of seeing.

But for our purposes, the reason I present this trichotomy is to draw attention to approach 3. It's a playing out of the pattern- burying texts by praising them. Don't be so finicky, don't miss the forest for the trees, It almost would be an insult to the text to be so very jejune as to worry overmuch about whether the formulas it presents for determining value etc. are quantitatively accurate. Don't you know that it's a critique of political economy, not just another work of political economy in its own right? The degree to which these interpretations of Marx conform to what Marx actually thought isn't our concern here. Rather our concern is that these arguments let the theorist get out of having to defend or attack Marx on these points not by diminishing the authority of the text, rather the opposite, by making it above such mundanities as what it appears to say at its face. The method is very different to Rabbinic Judaism- a focus on conceptual abstraction- both of methodology and content- more than a superabundance of meaning. Despite these differences, the result is similar, the text is made to have greater authority by the same intellectual maneuver that makes the literal meaning matter less.

VII.Example: Catholic Christianity

The method by which Catholicism overcomes literalism via veneration is through salvific contextualism. The bible is holy because it is part of God's holy plan for the salvation of humanity. As part of an infinitely subtle plan, it would be a mistake to see it as a go-to "grab bag" of readily available truths. The book is made more holy (through integration into an incredible cosmic plan) at the same time as it is made less literal.

Catholics believe that the bible is infallible in moral matters, and matters essential for salvation (though not on most matters of empirical fact):

"The books of Scripture must be acknowledged as teaching solidly, faithfully and without error that truth which God wanted put into sacred writings for the sake of salvation."

Nonetheless, Catholics will tell you that the bible, on its own, can't be read as a guide to doctrine. Catholics will generally cheerfully admit that by reading the bible alone earnestly, you could arrive at a form of Protestantism, but they will continue that, reading the bible earnestly you could arrive at almost anything. Thus, maybe, sincere individual reading isn't up to the task.

It may be instructive consider what has generally happened to Christians who insist that the bible alone should structure Christian faith. There are at least 200 protestant denominations in the United States- and this is a conservative estimate.

The Protestant conceit, say Catholics that we're each meant to pick up this book, well over three quarters of the way to a million words, and each study it as if for some private final theological exam by the Lord God is gives the Lord no credit, as if in his providence he forgot to provide teaching for us. Trying to read it alone for the purpose of finding salvation is a bit like setting out to cross the Atlantic in a rowboat. What kind

of cruel God would setup up a 700,000 word puzzle box, and only if you get the right answer do you get to go to heaven?

As a result, many Catholics will say- only half jokingly- that you shouldn't really be reading the bible alone unless you have a theological degree. "Only priests, rabbis and experienced monks and nuns and scholars should be permitted to read the bible" said one Catholic I follow on Twitter.

Sometimes individual Catholics will- more out of a little irreverent humor than theological conviction- disparage the bible when making this case.

There's a kind of riposte that you can make to this line of reasoning. Take some set of Catholic beliefs that can't really be found in the scriptures, say the perpetual virginity of Mary. The problem is not that just that it isn't obvious in the text- that could be dealt with easily. No, the problem is that the perpetual virginity of Mary is apparently contradicted by the text. You can make a pretty good scriptural case against it using many bible verses like:

"Isn't this the carpenter? Isn't this Mary's son and the brother of James, Joseph, Judas and Simon? Aren't his sisters here with us?' Mark 6:3"

Of course you can dispute this. You can claim, as do many Catholics, that the word translated here "brother" can also mean "cousin" in ancient Greek. I am told scholars of ancient Greek generally do not find this persuasive. Or you can go the Eastern orthodox route and say that these were children of Joseph by a previous marriage. I'm sure there are other possible lines of argument too. Who knows, maybe one of them is right, but even if they are, that still leaves the question why did the Lord God choose to make his book so misleading on matters of doctrine? Leaving things out would be one thing, but including things that, on a superficial reading, contradict his holy church seems like something God wouldn't do.

Monsignor Bransfield has an interesting response to some related concerns to this, that I think exemplifies some of the moves any Catholic is going to have to play:

Some beliefs are more hidden. Love loves to hide secrets, so that when we find them we are enraptured even more by their beauty. The mystery of Jesus is so profound that sometimes you have to look closely to see all the parts that he has made known. The Holy Spirit has hidden some dimensions of the mission of Jesus in the Bible. The truths of faith are clarified by the Tradition through the Magisterium, the Church's authentic teaching office. These truths never contradict the Word of God in Scripture, but serve to articulate its truth more clearly.

What Monsignor Bransfield has begun to do here is conceptualize the bible as something other than a document containing a collection of things you should believe. The bible isn't meant to be a succinct declaration of good teaching, it's meant to be a challenging document that serves many different purposes for many different audiences. In parts at least it is meant to be confusing.

The reasoning makes at least some sense. God is omniscient and infinitely intelligent. He anticipated every single person who would ever read the bible and what their response to it would be, over thousands of years of history. For those billions of people he had billions of purposes, to teach yes, but also to chastise, improve, test, inspire awe and humiliate. It is an infinitely planned document.

And for exactly that reason, you can't just pick it up, grab the most natural meaning that occurs to you, and read that as true doctrine. The process is perhaps most similar to the Jewish case, though the emphasis is a little different, one focuses on deepening the text whereas the other focuses on expanding the links between the text and a broader salvific context.

Once again, the very sacredness of the document becomes a reason you can't take what it appears to say too seriously. Once again, meaning is trapped behind a gilded cage of respect. I'm not saying the thought process isn't justified, but that's the result.

VIII. Example: American secular religion and the constitution

The method by which the American Civic religion venerates its holy book- the constitution- while at the same time evading literalism about it might be called synecdocheism. The text is transformed from a text into a symbol meaning something like "the wisdom of our founding sages". As the text is no longer primarily a text, but a holy symbol, it simultaneously becomes more divine even as its literal meaning fades.

The American constitution, interpreted literally and by original intent, suits the interests of no one in America except maybe a very particular type of libertarian (and even then...). Liberals, to their credit, at least acknowledge this implicitly with their "living constitution" theory, that the constitution develops over time, although conservatives respond, not implausibly, that this is just a form of Kritarchy. The conservatives haven't really got an alternative though.

Despite disagreeing with much of it, all patriotic Americans love their constitution. Let's review some contradictions:

While the practice of keeping an enormous long term standard army is probably legal in the letter of the constitution, it is at least very arguably against the spirit of it: "The Congress shall have Power To . . . raise and support Armies, but no Appropriation of Money to that Use shall be for a longer Term than two Years". Yet that doesn't bother anyone very much.

Even those who defend the electoral college, seem to have very little interest in implementing the electoral college as it was intended- a body of educated men

appointed by the states who have a discussion with each other and choose who should be president.

I would argue that a respectful approach to the constitution has no place for the qualified immunity doctrine, which makes American government actors often pretty lawless in relation to the constitution. The qualified immunity doctrine is (theoretically) a creation out of statutory interpretation, yet it utterly blunts the effectiveness of constitutional protections. Yet while both liberal justices and conservative justices have hemmed and hawed about it, they haven't gotten rid of it.

It looks possible that for the right case the supreme court justices might be willing to substantially reform qualified immunity, but even if it is reformed in the future, the justices seem in no great hurry. Outside of the courts a majority of conservatives, the faction which loves the constitution the most, oppose eliminating qualified immunity because support for qualified immunity has become coded as a pro-police issue.

Judges balance the plain meaning of the text against their policy preferences, and an assessment of what they think they can get away with. Whether conservative or liberal, they always have and they always will.

Yet despite Americans not taking the constitution literally, except where it suits them, they certainly take it seriously. There's a great Onion article entitled "Area man passionate defender of what he imagines constitution to be". American passion about the constitution isn't weakened by their loose regard for what it actually says, American constitutional veneration is enabled by this loose regard. It has (at least partially) moved from being a document that says specific things, which one might agree with or disagree with, to being a synecdoche of all things that are good in American life. Though they have used a very different method to the Jewish tradition's relation to the Torah, the Marxist relation to Capital or the Catholic relation to the bible, they have

come to a similar result. They venerate the constitution not in spite of, but by means of, the very same process through which its specific meaning becomes secondary.

The only difference is that the process is a lot simpler here. To the extent it is thought about at all, it goes like this, taking the example of qualified immunity.

Axiom: Constitution=Good

Axiom: Pro Police=Good

Lemma: Constitution=Pro police

Axiom: Qualified immunity=Pro police

Theorem: Qualified immunity=Constitution

Americans avoid the literal meaning in a very different way, they almost don't think about the meaning at all except where it suits them. Rather they transform the text into a symbol. Because it is a symbol, not a text, it almost doesn't have a meaning anymore- at least not in the way texts have meanings. The result is that, again, veneration reduces literalism.

The Mystery of Love

I wanted to write something about the lyrics of my favorite song The Mystery of Love by Sufjan Stevens. The song is most famous for being in the film "Call Me By Your Name". Personally, I think that's a shame because I think it's a song that deserves recognition, and I worry that being tied to a context, especially a controversial one, will prevent that.

I've copped shit for liking this song. It's often considered low-rent among Sufjan Stevens fans, partly because it was such a breakout hit for Sufjan- the bit of Sufjan that

non-specialists know and like. It's like saying you like a band, being asked what songs you like, and listing the first songs on the greatest hits album. But when the crowds are right, they're right. This is a song of power.

Someone told me there's a cultural movement of people who want to live in a way inspired by *Call Me by Your Name*. I have mixed feelings about that, I'll just say that I hope they're taking the right lessons from it. I question the wisdom of trying to model your life on a single piece of art. However, if you did want to look to art for guidance in living, you could do much worse than *The Mystery of Love*.

Most readings I've seen of *The Mystery* try to explain it in terms of the movie "*Call Me By Your Name*". While parts of the song deliberately parallel the film, in the main I don't think this is the right way to approach it. I think *The Mystery* is a song all about Sufjan, and the song's integration with the movie is at relatively shallow levels of meaning. It reaches the same conclusion as the film- a conclusion we will return to "it is better to have loved and lost..." but it does it in a more Sufjan way.

None of this is to say that the song is necessarily an honest look at Sufjan's lifeworld. It could very easily be wholly duplicitous. But treating the song as being about Elio and Oliver is only going to get us the top layer of the song. To dig deeper into the tell is to dig deeper into Sufjan's personal mythology.

The first thing to understand about the song is that three of its characters- Rogue River, the narrator's lover, and God become symbolically identified with each other. The song is very clear about this identification- it's not something you have to strain to see. Once you do see it, the song begins to disclose itself to you.

Read through the lyrics online, and then we'll go through them verse by verse.

Oh, to see without my eyes

The first time that you kissed me
 Boundless by the time I cried
 I built your walls around me
 White noise, what an awful sound
 Fumbling by Rogue River
 Feel my feet above the ground
 Hand of God, deliver me

The motif the song will keep returning to is opposed statements following each other.:

“To see” but “without my eyes”.

The narrator is “Boundless” and yet he has “built your walls around me”,

The contrasting “awful sound”- with the ecstasy of fooling around by the river,

A feeling of joyous flight “feel my feet above the ground”, contrasted with a prayer for deliverance.

What’s being seeded here is a unity of opposites, a unity that will bloom later in the song.

Note that, in the first verse, the beloved, the river, and God have all been introduced.

Oh, oh woe-oh-woah is me
 The first time that you touched me
 Oh, will wonders ever cease?
 Blessed be the mystery of love

Lord, I no longer believe

Drowned in living waters
 Cursed by the love that I received
 From my brother's daughter
 Like Hephaestion, who died
 Alexander's lover
 Now my riverbed has dried
 Shall I find no other?

The contradictions speed up, like tumbling waters, they come in the same line.

“Lord I no longer believe” (why invoke the Lord if you do not believe in him?)

“Drowned in living waters” (How do the waters of life kill?)

“Cursed by the love that I received” (How can the love of a child curse? Answer: this is a song not so much of the destruction of innocence, as a song of the destruction by innocence.)

Now we come to perhaps the biggest contradiction of the song and one of its most famous lines. In what sense is the speaker “like Hephaestion who died, Alexander’s lover”? Hephaestion is surely the opposite of the narrator for Hephaestion died, but he kept his love, whereas the speaker lived, but lost his love.

There is an implicit resolution to the contradiction- dying and losing such love is equivalent. It’s pretty grim and presages the narrator calling upon the river (as a representation of his lover) to kill him.

“Now my riverbed has dried”. Yet another contradiction, for just moments ago the narrator had been drowned in a flood of living water. There’s something fascinating about the dream logic that holds the flood and the drought to be one. I don’t fully

understand it, but as I sat down to write this, it occurred to me that the use of multiple and contradictory metaphorical vessels to try and capture one despair can be a way of saying, intentionally or not, “this despair, and the mystery of love which gives birth to it, exceeds any particular metaphor”.

Oh, oh woe-oh-woah is me
 I'm running like a plover
 Now I'm prone to misery
 The birthmark on your shoulder reminds me

There's a strange backwardness about the line “The birthmark on your shoulder reminds me”. Presumably, our hero won't be seeing any birthmarks on his beloved's shoulder anymore, so how can the birthmark remind him? Shouldn't he say that he is reminded of the birthmark on his beloved's shoulder rather than being reminded by it? Let me indulge in a tremendous act of critical speculation on a very narrow ledge of evidence. The key to understanding this is that time is very loose in the song. He is even now still with his beloved by Rogue river, and even when he was with his lover by Rogue river, he was already mourning his lost love. Because all is of the one nature, time is irrelevant. A corollary of this is that, at least in the emotional logic of the song, to reject the mourning of the present would be to reject the glory of past love. They are one. The narrator will later come to this conclusion himself, concluding the song by blessing the mystery of love.

How much sorrow can I take?
 Blackbird on my shoulder
 And what difference does it make
 When this love is over?

This nails something commonplace very well- “What difference does X make now that Y” is perhaps the prototypical grief thought.

I confess I don't really have an understanding of the line about the blackbird. It may be a reference to a line in the CMBYN book (not represented in the film), but this is just parallel spotting, it doesn't really help us understand it. Although I'm pretty certain it's not what Sufjan had in mind, I am reminded in a roundabout way of the ravens Huginn and Munin, perching on Odin's shoulder, representing Odin's intellect, and thus wisdom purchased at a terrible price. Odin gave first his eye, then hung himself on the world tree, symbolically dying and rising from the dead nine days and nights later.

Shall I sleep within your bed?

River of unhappiness

Hold your hands upon my head

Till I breathe my last breath

Oh, oh woe-oh-woah is me

The last time that you touched me

Oh, will wonders ever cease?

Blessed be the mystery of love

The poet ends by blessing love, the mysterious river of gain and loss, living and drowning waters. Helpless to steer it, or even protect himself from it, his only choice is how to regard it, and, like throwing a gold coin into an ocean storm, he blesses it. Death and life, love and loss are one stream. That stream, despite its self-contradictory nature, must be blessed, as a whole.

I'd like to think that, at the end, the narrator has begun to find the path out of his suicidal despair. The poem brings to mind one of the most popular "thought-terminating clichés" about gratitude for a lost love from Alfred Tennyson:

“I hold it true, whatever befall... Tis better to have loved and lost than never to have loved at all”

Sufjan arrives at the same conclusion but spells out the metaphysics a little more. The joy and the grief are one. The beginning, the middle and the end of the song and the love are united by one refrain “Oh, oh woe-oh-woah is me”. They are united by one river that runs through the whole song. They are united by one blessing, for one Mystery- the Mystery of Love. All are part of a glorious totality that we participate in.

To see the grief at the end as evil is to see it only very partially, for it is part of a greater whole, love. Who among us would dare to call that whole evil? If this doesn't quite make a peace treaty with grief, it at least sets that table for negotiations. I don't fully agree with this metaphysics, but I am captivated by it. If you want something to live by, something that will give you the clarity to be wise and the joy to love and fight, take it from this song.

Both Tennyson & Sufjan are poets of grief, homoeroticism, and homoerotic grief. These things have fascinated poets and their readers, because grief is unsayable, and the homoerotic has traditionally been unsayable and is still difficult to talk about because we do not yet have the same stock house of tropes. A poet is someone who wants to say the unsayable, and a mystic like Sufjan all the more so.

Now you may disagree, but I would hold “It is better to have loved and lost” and “blessed be the mystery of love” have something else in common in that they are both statements of will. They are not an observation about the world, rather they are a determination by the author to resolve things one way rather than the other. Sufjan has decided to bless the mystery of love (a manifestation of the God of the song) an action more impossibly arrogant than anything in the world except its only alternative, cursing the mystery of love, and hence God.

The river, Sufjan's beloved and the Lord, merge into each other, then they are all accused of drowning him in living waters. The theology is pantheistic. But it is not pantheistic in a lazy, modern way where pantheism is a cover for a more "spiritual" atheism. Rather, pantheism is something more properly called Panentheism or Theopanism. God is the world, yes, but he is also very much transcendent from it. This view of God makes the problem of evil- how an all-powerful and all-benevolent God can exist in the same universe- all the more urgent. If God and the universe are not distinct, and evil exists in the universe, does this not entail that evil exists within God? If evil exists within it God, doesn't this mean God is, at least partly, evil?

The narrator is not interested in answering this question, at least not explicitly. He poses the question sharply and gives his blessing to the mystery of love- in both joy and horror. He then offers his body to be taken by the waters.

This is not the most lyrically original song, let alone poem, in the world. It certainly is not the most thematically original. However, the poem has sincerity and simplicity fitting to its theme. Its echo of other romantic poems and tropes works to its advantage. It is more sentimental than cognitive, and this also works to its advantage. It retells one of our most fundamental stories- a lover is driven to existential questions and contemplation of suicide by the loss of a beloved, but it tells it so well, with focused artistic skill from a poet old enough to have mastered technique, but still young enough to have his full powers.

It is both my favorite song and poem, in the world.

The real climax of Call me by Your Name

This is an old mini-essay of mine that I thought I'd append, as it's on a related topic

I watched CMBYN for the first time on Sunday, and I have a theory about the real climax of the film. Maybe it's a bit pretentious for someone who has seen the film once to share a theory in a world of superfans, but here goes.

As we've all noticed, the question "is it better to speak or die" (with an interesting parallel to the traditional invocation at weddings- "Speak now or forever hold your breath") reoccurs several times throughout the film. These words are even more significant than the words that make up the title.

During his call to Elio, Oliver asks Elio if he minds Oliver's upcoming marriage. It would be easy to think that this is a throwaway line, a courtesy, but I don't think it is. Every word is significant in this part of the film, I don't think it would be included if it were a mere courtesy. I think Oliver is hoping that Elio will say something to save Oliver from himself. In fact, I think these words, and the silence that follows them, are the real climax of the film.

In the script Elio says "you're being silly"- with a double meaning- he could mean "it's silly for you to get married" or he could mean "you're silly to even think I might mind". Through ambiguity, he carefully avoids really saying anything. In the film itself, unless he mutters something so low that I can't hear it, he doesn't respond at all he looks like he might be going to respond, then his parents pick up the phone.

Elio is faced with a choice between speech and death. He chooses a metaphorical death.

It's pretty obvious in the film that Oliver's lack of courage- his inability to live in a way that is authentic to his same-sex desires- dooms his relationship with Elio. He chooses metaphorical death over metaphorical speech. What is perhaps less obvious is that, in this moment, Elio joins him in that indecision, and also chooses death over speech. Both are the knight, both are the princess, both choose to hide what they feel rather

than reveal it. Both lose a part of themselves. It's not fair of course, Elio's lapse is much smaller than Oliver's- a mere moment of indecision, but life rarely is fair.

I think this is the meaning of one line of "The Mystery of Love" White noise, what an awful sound- sometimes there is no middle ground between silence and making a noise.

Postscript: Sometimes it is better to die

I realized, rereading the above essay, that I probably gave a misleading impression. I don't necessarily think that in choosing to "die" Elio made the wrong choice. He is under no obligation to plead for Oliver to come back to him and abandon his fiancée, and in many ways, this would be a risky and perhaps even dishonorable option.

Outside romantic tragedy and comedy in the real world, it's sometimes better to hold your silence and never speak. We don't see much of this in art, because it doesn't make a great story, but that's the world we find ourselves in. Speak or Die is so compelling as an invocation, exactly because often there is a good case for both.

The problem of simulator evil

I've been going through Chalmers's book *Reality+*. It's a good refresher on some of the more interesting implications of simulation theory. I noticed that he'd come to many similar conclusions to me on a variety of topics, so I figured I'd best get what remains of my thinking on these topics into print as quickly as possible before he does so that I can claim priority ;-).

In particular, I wanted to hone in on a question- a kind of modern update on the problem of evil. If we are in a simulation, does it follow our simulators are bad people?

A brief summary of the argument we're in a simulation

Readers who are already aware of the simulation argument can skip this

Why think we might be in a simulation? This is my version of the argument, which draws elements from both Bostrom & Chalmers. It's a little closer to Bostrom than Chalmers because I find Bostrom's version more persuasive for reasons I won't get into here. My version of the argument is not as technically complete or comprehensive as it could be, because it is designed to be accessible. Nonetheless, it is, I think, in essence, right, at least on the basis of the evidence available to us at the moment.

What it "feels like" to be in a simulation is the same as what it feels like to be outside a simulation. Two people in the same situation (but one simulated) with the same past (but one simulated) will have the exact same experiences.

If humans survive the next few hundred years (at the most), human nature being what it is, it seems likely we will create many simulations, including simulations of humans. These will include simulations of our past- before we gained the capacity to create detailed simulations. Call these "ancestor simulations".

The capacity to create simulations is abundant- potential computational power is vast. Our curiosity and desire for entertainment is also abundant. It is therefore likely that, if we start creating ancestor simulations, we will create a vast number of such simulations of our history, many times the number of simulated people than the number of people who ever existed.

Since by (1) we have no other evidence that would discriminate whether we are in a simulation, we need to fall back on the baseline probabilities.

By (2 & 3) the baseline probability that we are in a simulation is higher than the baseline probability that we are not in a simulation,

Ergo we are probably in a simulation.

Chalmers on the case that our simulators are divine

As Chalmers notes, simulation theory. has been called the most interesting new argument for theism of modern times. If we are in a simulation, then our simulators are:

- Our creators
- Enormously powerful with respect to us.
- Have at least the capacity to be enormously knowledgeable about our lives, even if they don't choose to exercise it.

These features can be seen as corresponding to traditional divine attributes. God(s) are generally thought to be creators and immensely powerful. Many, though not all, traditions hold that God(s) know all things or at least a vast amount. Thus the simulation argument can be seen as generating a kind of limited theism.

Our simulators have other interesting features as well in this regard- for example, being outside time and space with respect to our simulation, corresponding to Boethian concepts of deity.

The problem of simulator theodicy

But there's another divine attribute, particularly important in the Abrahamic religions (though not only those), the attribute of omnibenevolence. It's far from clear that if the simulation argument is true, our simulators are omnibenevolent. In fact, you might worry they are evil- or perhaps somehow beyond good and evil (which is to say, in practical terms, evil). There are two arguments one might use to derive the conclusion that our simulators are evil:

The argument from suffering (and the absence of bliss). This world is filled with suffering. A good simulator would not create beings that suffer and would create beings that experience more bliss than us. Note that this can be extended to other evils besides suffering- for example, a lack of freedom.

The argument from deception, a good simulator would not deceive.

Our question then is: suppose our world is a simulation. Is the way the world is compatible with our simulators being good people who have made the world this way deliberately?

By good person, I don't necessarily mean anything particularly demanding. Certainly not omnibenevolent. Perhaps the best definition of what I mean in this context is:

A good person is a person who does not cause substantial harm to others without a justification strong enough to excuse that harm.

A lot of this is going to come down to divergent values. My personal sense is that the argument from deception is relatively weak- *ceteris paribus* our simulators would owe us the knowledge we are in a simulation, but even a relatively modest justification could get them off the hook for not telling us we're in a simulation.

Thus we'll focus on the argument from suffering (and other evils).

This is not just an abstract philosophical question. Though we probably cannot do much about it, it is possible that no question matters more. Our simulator could well be omnipotent with respect to us. They could turn us off, create natural disasters, wipe some of us from history, send us to heaven or send us to hell.

Does our simulator owe us any more than a greater than even lifetime balance of good over bad?

One of the best defenses of our simulator's moral goodness is to try and lower the bar for goodness as low as possible.

We should take seriously the idea that perhaps all our simulators owe us is more good than evil across our lifespan. One could even lower it further, and argue that all they owe us is for humanity as a whole to experience more good than evil across its lifespan.

Suppose you were speaking to your simulator. You had a dialogue with her reminiscent of Job- accusing her of badly mistreating you. To this she replied:

“Would you prefer you'd never existed?”

“No, but you could have made things so much better!”

“Yes, but I’m not running a simulation of paradise, I’m running a simulation to find out about something, and having all simulated beings in a state of perpetual bliss would interfere with that. Nonetheless, I’ve taken steps to ensure that all lives in my simulation are worth living [ed: this could be achieved by running only a sparse simulation of the most miserable lives, or perhaps through a simulated afterlife for those who found earthly life worse than not existing at all] Or at the very least I have taken steps to ensure the total experience of the simulated human species is more positive than negative. I get the data I want. You get lives that are worth living- either individually or at least in the aggregate. in what sense can I be said to have wronged you?”

“You could easily make things better, but you choose not to, that’s wrong.”

“I can’t make things better easily. I have a limited computational budget for simulations.”

“Why aren’t you spending your computational budget on creating blissful lives?”

“This simulation is being run for some kind of purpose in my world- perhaps science, perhaps even entertainment- I won’t get into the details. I have the budget I do contingently on meeting that goal. If I just created blissful lives my funding would be taken away. Thus your choices are non-existence or the lives I give you. On the whole, I think this benefits both of us, and doesn’t make me evil ”

Whether this is an adequate response is going to depend on your ethical views.

However, I think it’s clear that there is at least a coherent conception of the good on which what our simulator does in this scenario is defensible. Thus we can’t be sure that our simulator is malign.

Is it immoral to switch off a world, or a person at death? This depends on whether death is harmful.

One of the more terrifying implications of the simulation hypothesis is the possibility that the simulator could turn it off at any time. An interesting question then is if our simulators are benign are they be obliged not to turn us off? At least without our consent?

There is an ancient debate in philosophy over whether or not death is a kind of harm. That is to say, if someone dies, is that, in and of itself, harmful for them? The answer to this question will establish whether or not our simulators could count as benign, and still turn us off. Epicurus, for example, thought that death was not harmful.

This, I think, is just going to come down to personal intuitions on death and harm. I won't go through the philosophical arguments here. My sense is that the majority of people if they thought carefully about it, would come to the conclusion that dying is bad.

If our simulators are benign and regard involuntary death as harmful, this has interesting implications beyond the question of whether they can turn the world off as a whole. It would tend to suggest that we could expect that death is not the end, and the dead are spirited away to some sort of afterlife- although perhaps death is, while tragic, necessary for some reason in a way that justifies our simulators allowing it.

Even if death is not intrinsically harmful, it might be held that dying after an unsatisfactory life that you would be better off never having lived is a sort of harm. Simulators might have a special duty to correct this through an afterlife. A similar argument might be made about premature death- although what counts as "premature" from the point of view of a god-like simulator might be difficult to assess.

Can we know that the various evils we complain about exist?

One thing we need to consider is that if we are in a simulation, our evidential basis for judging our creator is sketchy. Granted, the epistemological and metaphysical issues are complex, as Chalmers discusses, but it seems to me that if we're in a simulation we can't be confident that the past of that simulation happened the way it appears to have happened.

Any given awful experience that you might hold against your simulator might have never actually happened. The scope of evils for which the simulator is responsible might be far smaller than it initially seemed (or larger!)

Even the basis of our reasoning is suspect. It could be that inferences that appear plausible to us are the result of manipulation by our simulator. How easy would it be to manipulate us so that we all think $2+2=4$, when really it equals five?

These kinds of skeptical doubts start tearing up the very bases on which we came to. This leads to an argument that skepticism is self-undermining.

I do tend to think that, at a certain point, skeptical doubts become self-undermining, but theorizing exactly where this point is, is difficult. Chalmers quotes one of my favorite philosophical arguments by a physicist, Sean Carroll's argument that the idea we are Boltzmann brains {one of the most extreme skeptical hypotheses} is self-defeating- I tend to agree with Carroll on this. On the other hand, I'm sure that some philosophers will try to argue that the idea we are in a simulation undermines any evidence we might present for it, but I find this implausible.

The truth of where to draw a line against doubts as futile and self-undermining probably lies somewhere between Boltzmann brain and ordinary simulationism. In our inquiry into the moral character of our simulators, I see little option but to proceed on the basis

that, while our world may be simulated, things happen in the simulation as they appear to while expanding the error bars around our conclusions.

What if we live in an ethically driven project- Diversity Utilitarianism

Another possibility that we need to consider is that if we are in a simulation, we may be in an ethically driven project. By “ethically driven project” I mean a project that exists for our own good, and/or the good of humanity. So long as our simulators have similar ethical values to us (a big if) this would be a fantastic outcome. There are many different possible ethical projects we could be a part of, in the next two sections I’ll consider two of them.

Why would they put us through pain and suffering if they are working for our own good?

Suppose I gave you vast, though not unlimited, computing power and put you in an otherwise empty universe, what would you do? If you’re anything like me, you’d want to create numerous beings, and let them live blissful lives. Perhaps humans, because we’re biased.

You might also feel like these beings have to be genuinely distinct from each other, and live varied lives. A vast number of copies of a being experiencing a single blissful moment over and over would be unsatisfactory.

Call this position diversity utilitarianism. A diversity utilitarian holds that total value is equal to the sum of the utility of individuals. However, this value is diversity weighted in some way. If there are two beings, Don & Nod, and they are quite distinct from each other, total utility equals the sum of their utilities. If they are identical, total value is maybe equal to half their total happiness, or perhaps just a little over half their total value. If they are very similar, but not identical, perhaps there is some penalty to how much their aggregated utility is worth.

Personally, I find diversity utilitarianism plausible, at least in so far as tiling the universe with identical simulated people experiencing bliss doesn't sound that attractive. If our simulator is a diversity utilitarian- or something similar- they will need to generate not just as much bliss as possible, but diverse bliss.

How do you create numerous different humans, genuinely distinct from each other? Well, it's possible that the most efficient way, or possibly even the only feasible way, to create a human personality - especially a range of different personalities- is to simulate the biological and social processes of human life. Our world could thus be a diversity utilitarian people generating ground.

But why not generate these future citizens of blisstopia in a blissful world? If you want the humans you create to be diverse, just raise them in diverse blissful worlds. Chekov said that all happy families are the same, it's the unhappy ones that are different, but surely Chekov aside, there are uncountable possible utopias.

I grant that, if you're motivated by the ethical goal of increasing total human flourishing, you'd start by creating blissful lives. But a posthuman civilization might have vast computational power- so much that they could simulate all sufficiently psychologically distinct beings that grew up in blissful conditions. Thus they might turn to simulating people who grew up in less than blissful conditions. After they died, or at a certain age, or something, you'd harvest them out of the simulation and set them up in a nice afterlife.

In other words, if this speculation is correct, we are the product of an attempt to balance psychological diversity with psychological bliss, after the low-hanging fruit of people raised in utopias has been exhausted.

That scenario probably sounds absurd, or wishful thinking, but it first occurred to me not when thinking about this problem, but when thinking about what I'd do if you gave me vast computational power. It has a degree of independent plausibility.

What if we live in an ethically driven project- Nikolai Fyodorovich Fyodorovism

Nikolai Fyodorovich Fyodorov is my favorite non-Marxist Russian philosopher. Nikolai believed that the greatest source of alienation in our lives is the alienation of the living from the dead. We are cut off from ancestors and friends alike by that dread scythe. Nikolai, however, had a can-do attitude. Where a lesser, perhaps saner, philosopher would simply bemoan the tragedy of death, he proposed its abolition. But he went beyond the normal transhuman desire to eliminate death- for he wanted to eliminate it retrospectively. Nikolai wanted to raise everyone who had ever died from the death.

Another reason you might simulate people with less than blissful lives is if you wanted to complete Nikolai Fyodorov's universal resurrection project. You wanted to recreate every human that had ever lived because you thought you had a duty to resurrect the dead. Since historical information is partial, in order to be sure of creating a good psychological approximation of everyone, you'd have to make a vast array of attempts. Certainly, there is enough mass and energy for a vast number of attempts, although just how many is a little unclear.

And so, on this theodicy, the bad stuff we experience is in a strange sense, formative. It is necessary to bring us back into being.

Now you might be wondering "couldn't they just skip the experiences and create people without actually simulating the life history?" The answer may very well be no. It could be that there is no way- or at least no computationally efficient way- of creating personalities without running through a simulation of that personality's history

The problem of quantitative theodicy

Scott Alexander presents a kind of Theodicy that converges with what we called diversity utilitarianism but in a non-simulator context. Essentially, God aims to create as much (net) good as possible. First God creates all possible completely good worlds, and then when he runs out he creates worlds that have some good and some evil in them.

This makes me wonder. Chalmers claims that there is enough capacity in a kilogram of matter to simulate 100 years of life for 10 billion people. The mass of the galaxy is 1.5 trillion solar masses, which I think is about 10^{40} kilograms. Is it plausible that using the mass of the galaxy to create simulations, one would run out of diverse, blissful lives, and have to resort to mixed lives like our own?

Now theodicy is reduced to a strange sort of maths problem, albeit an insoluble one, since we do not have any quantitative sense of how much diversity is required, or a way to quantify diversity.

We also don't really know how much matter our simulators have. Perhaps they have far more than a galaxy's worth, perhaps they have far less.

Consent theodicy- the virtual contract

Years ago I outlined a consent theodicy. I argued that it's possible that we consented to live in a world with evil, or that our creator knew that in the counter-factual in which we were asked "do you want to live in this world" and the full reasons we were living in this world were given, we would say yes. Hence we suffer evil because we have agreed to it? Why? Well, perhaps because it's essential for our development in some respect.

Obviously, such a consent theodicy can be combined with sim-theism. It is possible that you are in a simulation right now that you agreed to be in*. Alternatively, it is also

possible that your simulator would justify their treatment of you on the counterfactual that if you understood the full situation you would consent to be in the simulation.

*- [although this raises prickly questions about in what sense the person who agreed to be in the simulation really is you, I think there are plausible permutations of the conditions on which this turns out to be true]

Evidential decision theory and the simulation hypothesis- or why there's at least a modest case you shouldn't mistreat sims

Does our consideration of simulator theodicy have any practical implications? Well an argument can be made that it gives us reason not to create simulations maliciously, or mistreat them.

Quoting Wikipedia, evidential decision theory holds that:

The best action is the one which, conditional on one's having chosen it, gives one of the best expectations for the outcome.

Evidential decision theory is controversial. Its most prominent rival is causal decision theory, which holds that you should act in a way that is likely to cause the best outcome. Nonetheless, let's stick with evidential decision theory for the moment.

Now our world, as we see it, is compatible with a variety of simulators, some of them benign, some of them callously indifferent, some of them actively cruel.

It seems quite possible that our simulator is what we might term our value function descendant. A value function descendant of humanity is a being that has roughly our value function but is perhaps extrapolated out to remove inconsistencies and/or clarified. The argument for this is that, so long as malignant AI doesn't take over the planet, it

is likely that simulations we create and run will be run either by our value function descendants or by artificial intelligence under the control of our value function descendants.

Thus, if it turns out that we mistreat simulations in the simulations we create, the likelihood that we are in a simulation in which we are going to be mistreated goes up.

Therefore the action that gives the best expectations of outcome is not to mistreat any sims we create, because it's reasonably likely that our simulators have similar values to us. If we commit sim abuse, it's more likely our simulators are willing to commit sim abuse. Thus, according to evidential decision theory, we have a reason not to.

Excursus- if you think our simulators are either humans or the descendants of humans implanted with our values, our probable situation depends on a kind of ethics exam at the end of history

If our simulators are human or value function descendants of humans -and not aberrant or rogue actors but representatives of their civilizations-, then there's a sense in which our simulated humanity will get what it deserves. People like us choose our fate in an ethics exam at the end of history.

I've long wondered whether the evils of the world reflect mistakes or conflicts of interest. This is why I introduced the language of conflict versus mistake theory all those years ago. The answer of course is both but in a very subtle way, with malice and mistake interpenetrating in a dizzying web.

Suppose that, due to super-intelligent AI, we eliminated the possibility of mistakes. Do you have confidence that faced with genuine knowledge of the consequences of their actions, humans would choose to do the right thing? If yes, then rejoice because our

simulators are probably not malicious {assuming humanity is still in charge}. If not, then there's less comfort to be had.

What about the argument that even if humanity as a whole is good, we could have the misfortune to be in a simulation run by a rogue evil individual? It's possible but unlikely, I tend to think there would be a fraction as many such illegal simulations as legal ones.

Excursus- What would you do if you were powerful?

I think a useful exercise in pondering this stuff- not necessarily in arriving at truth, but in getting a sense of the dizzying scope of possibilities, is to consider what you would do if you were very powerful- say I gave you a billion dollars.

Having done that, consider what you would do if you were even more powerful- say I gave you the capacities of superman. What would you do if you were so mighty that you exceeded the power of all governments?

Now, having considered that lets up the power level again. Suppose that you were not just mightier than all governments, but also had a super-intelligent AI that would advise you on the best way to achieve your goals- whatever they were what would you do then? What values would you steer humanity towards?

Now we come to the highest pinnacle. What would you do if I gave you vast computing power- enough to create simulations of whatever you liked- and AI assistance in creating those simulations. What worlds would you create?

Excursus- Some broad value frameworks the omnipotent could have

Here's a smattering of different values systems simulators could subscribe to. Almost any of these value systems, in at least a partial form, can overlap with almost any of the others, and this isn't a formal classification, but it's a starting point for discussion.

For most of these value functions, I can imagine some possible way that our experience thus far could be compatible with a simulator holding this value function, but I'll leave thinking it through as an exercise to the reader.

Selfishness- hedonistic type: A simulator of the hedonistic type is dedicated to the satisfaction of their aesthetic, culinary, sensual, and/or sexual appetites. They may, for example, run numerous simulations to try and create the most exquisite and fascinating people to have sex with.

Selfishness- megalomaniac type: A selfish simulator of the megalomaniacal type wants to be worshipped, and to exercise their power according to their own strange whims for self-glorification.

Selfishness- aesthetic type: A selfish simulator of the aesthetic type views the whole universe as like an artwork of some sort.

Selfishness- scientific type: A selfish simulator of the scientific type is running the universe to answer some scientific question- regardless of whether it hurts the simulated.

Sadism: The worst possible scenario would be if we were in a simulation created by a sadist. This could come in several different forms- for example, they might be a general sadist, or they might be seeking revenge on a specific person or group, thus in the process of recreating them to torture.

Liberalism: A simulator of the liberal type wants to give us, above all, freedom of some kind. Exactly what that freedom amounts to will depend on the simulator.

'Crude' utilitarianism: A crude utilitarian simulator wants to maximize pleasure, or desire satisfaction or something like that, and so is running simulations to do so. We can be reasonably confident that we are not in such a simulation due to the existence of suffering.

'Diversity utilitarianism': As described above. A diversity utilitarianism wants to maximize utility - disutility. However, they weigh repetitive good experiences or good lives as worth less than non-repetitive good experiences or lives.

Humanism: A humanistic simulator sees its primary goal as the flourishing of people. It's a eudaemonist. Freedom and happiness, at least to some degree, are likely both parts of this goal, but neither is the full object. A humanistic simulator might need diversity for similar reasons the diversity utilitarian does- e.g. a flourishing life counts for less if it is a copy of an already existing one.

Fyodorovian: As above, a project to resurrect the dead.

Tribalism: A tribalist simulator is like a selfish one, but they dedicate themselves to a group, rather than just themselves. We are sadly not in the group.

Social Darwinism: A social Darwinist simulator wants to create strong creatures, for some value of strong, and even if it requires great suffering.

Primitivism: The primitivist singleton is leery of technology, and wishes to constrain it. This might sound like a bizarre or unlikely view for a simulator to take, but although I do not agree, I think it makes a certain sense. In the novel *Consider Phlebas*, by the sadly passed Iain Banks, the Iridians fight a war against *The Culture* because they view

the culture as devoid of human agency- AI does everything. Human striving and struggle is necessary for a meaningful existence, argue the Iridians. We can imagine a primitivist simulator who has put us in our world- just before the invention of artificial intelligence that can take over human functioning- for exactly this reason.

Moralism: A moralistic simulator wants to create good worlds, but their idea of goodness is laden with ideas that some might consider outmoded. Drugs are bad, promiscuity is bad, everyone must worship God, that sort of thing.

Radical aporia

I'd like to give a personal coda to all this simulation stuff, building on the brief discussion of skepticism, and branching out from there.

How are we meant to think about cosmology, and on a more personal level, the meaning and value of our lives in light of the simulation argument? We face both radical uncertainties about whether we are in a simulation and radical uncertainty about the implications if we are in. For example, what is the risk of being turned off? What does the future hold for us given that we don't know the purpose of the simulation? Does life end at death or do our simulators continue us on? If we are in a simulation, how can we be sure the past happened anything as we remember it, given that our simulators could just tweak our memories? But if we go down this road, how can we know anything about our situation, including the things that led us to posit we're probably a simulation in the first place? How can we even trust our own a priori reasoning, given that it would be trivial to interfere with that?

This all reminds me of Neurath's boat. As Neurath put it:

"We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way, by using the

old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction”

In truth, we've never known our own situation in the grander scheme of things. A lot of us thought that we had it figured out with a kind of vague, cosmological materialism, but we never had the full picture filled in on that story. There were always questions about the standard 19th-century materialist framework- the mystery of qualia (as Chalmers of all people has pushed), the Fermi paradox, etc.

Even the idea we are in a simulation only represents a guess given our current level of technology. Who knows what stuff we'll be pondering with the technology, social structure and speculative philosophy of the future? The simulation argument seems kind of persuasive with the tech of today, but perhaps the technology of tomorrow will suggest wholly different cosmic possibilities. To put it tautologically, we are conditioned by our conditions- things that seem like very good arguments to us now might seem like poor arguments in the future. Things that seem like poor arguments now, or that haven't even occurred to us, might seem compelling in the future.

In other words, I'm urging you to apply the skeptical meta-induction to speculative metaphysics. If it has power in the realm of science, how much more so in philosophy. Given how unstable our ideas have proven, not just about our cosmic situation, but even about what the possible alternatives are, we know nothing. We can't trust the simulation argument, can't trust the opposite, can't really trust anything.

So we don't know where we are in the logical space of possible worlds, not even approximately, as best I can tell we have no way of figuring it out. The only way to cope is to accept that you don't know, and you will very possibly never know, even the basics of your situation. Having accepted this, resolve to live by your values in a way that carries meaning even in an absurd and unknowable space of possibilities.

In an increasingly bizarre world, the thought that no one has ever proven it's not going to turn out alright can be a source of comfort. We're swimming over an abyss on a black night, and it's natural to worry a Levithan might be rushing up to devour us. That's possible, but hands might be rising up to cradle us as well.

Social commentary

Movements are always a distorted lens on the ideas they embody

Please don't hate me for this one. I don't think I'm better than other people. I'm just laying out in blunt terms what others have tiptoed around.

1. The problem

I want to spell something out in this post that I think many people know, but which goes under articulated because it's kind of mean. I apologize for this. I really do feel poorly about it. I don't like measuring people up in this way, but it's one of those things you've got to be sharp about to be honest about.

I remember reading an internal socialist party document once. It was very careful in its words, but essentially the document argued that during periods of "downturn" in class struggle socialist organizations which were "keeping the flame going" would inevitably attract dead-enders, no-hopers, people who were mostly looking for a social outlet etc. In the bible (1 Corinthians), St Paul remarks of the Christians of his time: "Brothers, consider the time of your calling: Not many of you were wise by human standards; not many were powerful; not many were of noble birth. But God chose the foolish things of the world to shame the wise; God chose the weak things of the world to shame the strong" i.e., the people joining up to Christianity at the time even by their own admission were not people conventionally seen as intelligent, financially successful or strong.

This isn't exclusive to tiny religious and political sects. It applies to pretty much all social movements. I don't want to be mean, so I won't name the movement but there is a certain social movement which I have in mind that regularly makes people remark "Urrrgh, X sounds cool in theory, but have you met the Xers? No thank you."

Bluntly, movements always have and will attract a lot of dead-enders. This is because

normal people don't need movements as an outlet—they already have friends and are capable of curating their own social lives. Dead enders are more likely to have trouble in this department, thus are more likely to gain something from the constructed social life offered by movements.

2. **N.B.**

Most of the examples I use here are from political movements because that's where I've spent a lot of my life. However I believe the dynamics I describe apply to a lot of other movements, from music (Punk) to even some philosophical movements (e.g. New Atheism)

3. **The taxonomy**

There are essentially four types of people who join relatively marginal social, religious and political movements, as well as certain types of intellectual and artistic movements:

1. People who aren't doing so well in the game of life.
2. Thoughtful, whip-smart, highly committed and charismatic people.
3. People intermediate between 1 & 2 with some features of both.
4. Genuinely normal people.

Category 4 is the rarest—so rare that we won't talk about it much further. Category 2 is the second rarest, more or less invariably. Whether 1 or 3 is the most common probably depends on what the movement is about and the context. I myself probably started life as a 1 and clawed my way, through the infinite patience of mentors, to a 3.

Why are these types particularly attracted to social movements? Well, it's no great mystery.

Category 1—The losers join social movements because:

A) They're lonely, and a social movement is a group of people who sort of have to be friends with you.

B) When you're behind anyway, why not gamble on something that might make it big? Become an early adopter of a plan to change the world?

Or the more sincere reason: C) Because their beliefs and values have been formed through a lifetime of exclusion from the dominant power structures. Thus their values and the values of critical social movements often have much in common.

To be clear, the reason the "losers" can't curate their own social lives is not always a lack of social skills. Sometimes, for example, they are perfectly charismatic but have burnt all their bridges through intermittent erratic behaviour. Sometimes they've just been unlucky in some way. A loser in this sense is someone who has turned to a social movement because they have to in order to cope. There's more than one path to that outcome.

Category 2—The stars are attracted to social movements for one of two reasons depending on the individual and how cynical they are

A) because it allows them to be a big fish in a small pond

B) because they're deeply committed to their beliefs, and will pursue them even if means hanging out with uncool people.

Category 3—The inbetweeners are attracted to these social movements for both sets of reasons in varying degrees.

4. The resultant neuroses

People in these groups, especially those who aren't dead-enders, are acutely aware of the dead-ender problem. The number of people who have started or led movements for a

time, only to lament that they can't stand their own followers, is huge. Often these laments are some variation on "they only understood the form of what I was teaching, not the essence." Rightly or wrongly, the dead enders are accused of not getting the ideas, or at least not the deeper underlying truths of those ideas. I don't know if it is actually true that the dead-enders don't really get the ideas, but it's a common perception. Certainly, some forms of dead-enderism seem to correlate with a lack of insight.

Sometimes people even go so far as to claim that the majority of people following the movement are actually a block on its success—that the movement would be better off leaner, but higher quality. My favourite contemporary example of sniping at one's own followers is the grandees of dirtbag left Twitter who are constantly complaining that their "reply guys" just ape them by repeating phrases like "normal country" and "Hellworld" without understanding the spirit of critical irony and convention busting that was meant to power it. Of course, this being Twitter & the Dirtbag Left, it's hard to be sure how much of the scorn is performative irony and how much is real exasperation—but my guess is "all of it is both ironic and heartfelt at the same time". There's something very funny about seeing would-be underground rebels reduced to complaining about slavish followers.

As has been remarked by 50 million other authors, the tension between the leaders and followers is most especially a problem for movements based around "breaking the rules". The leaders watch in despair as their modes of rule-breaking become the new rule by people who just don't get what it's really about.

I suspect these neuroses have been around for a lot longer than we realise. Histories of movements are mostly written about those we would classify as the "stars" of movements. Stars are also more likely to write the histories as well. This tends to submerge underneath the waters of Lethe lot of the angst about dead-enders. Still, nothing is entirely lost. It's been a while since I've read any of it, but I remember getting

the impression reading between the lines that some of these issues plagued 19th and early 20th-century socialism.

5. The contradictory role of size

Now you might think that the dynamics that I've described here only apply to radical ideas. They don't. I've met the youth wings and gone to branch meetings of all 2 1/2 major parties in my country if anything they were even a little bit sadder than the socialist groups, the libertarian groups etc.

Despite the fact that seemingly more mainstream movements are no less filled with losers, it does seem that as movements expand the people in them get more normal, and as they shrink they get less so. I remember Occupy in my home town, which continued for an unusually long period of time. It very clearly moved back and forth between these extremes(1).

David Graeber makes a similar point about anarchist and direct action groups in *Direct Action, an Ethnography*. He also points out that the proportion of women in these groups rises as the group grows larger, often closer in time to an important "action" or "event". I to have observed this, and I believe this gender dynamic to be linked, although I'm not going to feign a hypothesis about why it is so.

In *Ruling the Void* Peter Mair talks about the hollowing out of political civil society. Parties all over the world have fewer and fewer active participants. Based on accounts I have read of social movements throughout history, I do not think that this is isolated to formal political parties. Especially from 1990 to the GFC, activity was minimal. After the GFC there has been something of a slow recovery of activism, but it remains to be seen how permanent and significant it is. Activity in various movements is still modest compared with the 70's. This is not only true of political movements—even cultural movements seem pallid now. Sometimes it feels like only strange weirdos like myself remain.

6. The implications: evaluation

So why am I laying this out? Well, over the last five years or so, I've noticed a rise in arguments of the form "X is a bad idea because the people who follow X aren't living normal fulfilled lives". This takes a lot of forms, e.g. dueling Chad vs Virgin memes etc. Now I don't think we can dismiss this as an argument by simply labeling it ad hominem. If an idea is causing problems in living, and/or turning out and attracting unbalanced people, we should at least understand why before moving on. It is prima facie evidence that an idea doesn't work in practice if everyone who likes it isn't doing so well.

But I would caution against taking this argument too seriously, precisely because movements, especially their zealots, are always like this, and now maybe more so than ever.

The greatest potential for censorship is in the algorithm, not in the bans

I don't know a lot about Caitlin Johnstone, mostly I know her as a critic of United States foreign policy. What I do know is that she's right that most people are missing the real dangers of online censorship. She writes:

“But far, far more consequential than overt censorship of individuals is censorship by algorithm. No individual being silenced does as much real-world damage to free expression and free thought as the way ideas and information which aren't authorized by the powerful are being actively hidden from public view, while material which serves the interests of the powerful is the first thing they see in their search results.”

Let's expand a bit on this line of thinking.

The effects of algorithmic censorship

I'm going to start with a very brief sketch of what we know about where algorithmic censorship is up to.

We see when a platform bans someone. We generally don't see the changes that adjust how much we see of this or that. I'm not talking about shadowbanning here. Like banning, shadowbanning is a very vulgar and direct form of power. The most effective forms of censorship are more subtle.

Algorithms more complex than any single person can understand curate your feed. Before you see anything, these machine learning run algorithms consider the content of the post, who the writer is, who the people responding are, and many other factors.

We know that tech companies sometimes consider the political outlook of content when designing algorithms. Companies have specifically targeted political categories such as right and left. From a story in Mother Jones:

“Republican lobbyists in the DC [Facebook] office said, ‘Hold on, how will it affect Breitbart?’” recalls another ex-employee. Testing showed that the proposed changes would take a “huge chunk” out of Breitbart, Gateway Pundit, the Daily Wire, and the Daily Caller. There was “enormous pushback”. They freaked out and said, ‘We can’t do this.’”

The code was tweaked, and executives were given a new presentation showing less impact on these conservative sites and more harm to progressive-leaning publishers—including Mother Jones.

Thus the selection of what you see then is, at least sometimes, an explicitly political process. Political actors (lobbyists, even) help design the algorithm that determines what you see.

Other than that it happens, we don’t know very much about the explicitly political selection of content. Good business sense would be to keep it a closely guarded secret.

What we know much more about is the implicitly political selection of content. Tech companies, particularly Youtube, Facebook and Google, reduce access to sources they consider disreputable. They justified such changes by a perceived wave of misinformation after the rise of Trump, reaching its zenith with the coming of SARS-CoV-2.

Partly the exclusion of disreputable sources was achieved by directly targeting them. For example, Facebook tells page and group managers that it will penalize them for stories that have been rated poorly by “independent fact-checkers”. This can get truly absurd-

for example, once Facebook told me off for sharing something from The British Medical Journal that had been fact-checked- The British Medical Journal contested the fact check.

It's not just fact-checking though. Another part of trying to exclude "disreputable sources" is promoting results from mainstream media companies. The effect of shifting attention to mainstream media sources is to narrow the range of accessible political opinions. The political space of mainstream media has The Guardian on the leftmost pole, Fox News on the right-most pole, and The New York Times in the middle. The resultant political space is somewhat left of center by American standards and very right of center under a global standard.

I'm no fan of misinformation. I indulge in a smidgen of conspiracism (I don't think Epstein killed himself because he didn't), but I am mostly, uh, reality-aligned. I have sympathy with those worried about misinformation during a pandemic. However, we don't want the cure to be worse than the disease. It is not just bizarre claims about the vaccine & virus that are being targeted.

The quantitative effects of the changes on outsider media are huge. Consider, for example, the Socialist Equality Party. I've chosen them, in part, because I'm no apologist for the SEP. Their anti-union stance drives me nuts. They've got some very weird views on stuff that I'd really rather they didn't have weird views on, such as Roman Polanski. They're a minuscule, ineffective, sectarian Trotskyist grouping with outdated politics and troubling views on topics like sexual assault. Their analysis has little shade or nuance.

But, they do some good reporting on The World Socialist Website. They also sustain a good standard of factual accuracy- granted even by their political critics-, often ahead of that of comparable mainstream news organizations. Nonetheless, algorithmic changes

squelched their traffic by 67% in the 4 months between April 2017 and August 2017. Factual accuracy offered them no protection.

Google would say they didn't target the WSWS, rather they prioritized the establishment press. The establishment press, they would argue, is held to certain factual standards. Non-establishment press varies wildly in its quality. Google can't be expected to evaluate each website or article piecemeal. That's likely all true. However, I don't need to spell out for you the fact that the establishment press is not politically neutral. It is owned almost entirely by a tiny minority of people. Its crime reporters generally have close connections with the cops. Its foreign policy reporters are almost always embedded with the national security state. Its economic reporters are peas in a pod with the business community.

In other words, the mainstream press is an ideological apparatus whose separation from the state is often merely legal and formal. It acts as a mouthpiece for the establishment as a whole, or for sections of the establishment on issues that divide elites.

The character of algorithmic censorship contrasted with the character of normal censorship

Now that we have the factual background in place, it's time for the meat of the piece- The philosophy of censorship.

Our whole way of thinking about censorship- a critique that can be broadly identified with the 19th-century liberal philosopher John Stuart Mill- isn't prepared for content curation in the age of digital monopolies.

To talk about this sort of stuff rightly, you have to sound like a bit of a Foucauldian. Algorithmic censorship has features which mark it out from how we normally think

about censorship. Many of these features make me want to use words like biopolitical and panopticon.

I want to emphasize that these features are mostly not new in the history of censorship. They might seem new because a lot of effort is spent miseducating us on censorship. The algorithm applies the censorship of making the conversation rather than the censorship of excluding from the conversation. The censorship of making is not an unusual case in the history of censorship. I would conjecture, it is historically the most important form of censorship. We think of the censorship of exclusion as the most important form of censorship, because thinking of censorship that way is a pillar of liberal ideology. Let's rethink the core differences between excluding censorship and making censorship.

Fair warning: to better express the logic of a feed algorithm, I'm going to blur the line between where we are now, and where these trends could lead "in an ideal case".

Firstly, algorithmic censorship is inside the conversation, not merely around it: The algorithm is not a boundary keeping things out of the conversation. It is not a boundary corralling the conversation within certain parameters. Rather the algorithm censors content in the way the vascular system censors blood. Although some things are plainly disallowed, there is no such thing as merely "allowed". The algorithm assigns a portion of screen space from a subset of users to every bit of content.

Secondly, there is no ontological gap between censorship and conversation: It's not just that the algorithm is in the conversation. Rather, the algorithm has influenced generations of the conversation, rather like evolution by natural selection. Thus it has made the conversation. Creators that thrive under the algorithm rise to the top. The behavior of others is shaped by the algorithm, for two reasons:

1. They consciously or unconsciously model the successful creators

2. They see that their work is more successful when it is algorithm-friendly. This shapes their behavior like a rat pulling levers in a Skinner box.

In turn, the algorithms are shaped and reshaped as social media executives see what is good for the company. The line between the algorithm and the content is blurry at both ends.

Thirdly, algorithmic censorship operates in an ecological, rather than a regulatory manner, thinning and managing, rather than trying to eliminate: There is no pretension of eliminating unwanted views. A smattering can be quite useful. Such a smattering legitimates the site, and prevents dissidents from looking for kinder skies. This is particularly true if the algorithm can limit viewers of that smattering of dissidence to those specifically looking for it. In the ideal case no one who didn't already believe the dissident material would stumble across it.

Fourthly, the censorship is experienced as an aid to the conversation, rather than a hindrance: Censorship is not something that makes the discourse harder, on the contrary, it is the ground condition of being able to talk at all! Whereas other, more obvious, forms of censorship might have you losing your words, the algorithmic structuring of a media platform is experienced by most website visitors as an aid to finding the content they're looking for. Thus resentments don't accumulate, as they do against most censorship regimes, against the algorithm. Rather, resentment is focused at the iceberg tip- bans and shadowbans.

Fifthly, the censorship is "private" not "public": Some people are going to balk at this, maintaining that nothing can be true censorship unless it is public. Nonsense, consider the Hollywood blacklists during McCarthyism. These are enormous monopolies with intimate state connections we're talking about, not some boutique website selecting contributors.

This is not to say the state is absent- far from it. The censorship happens in a weird nexus of quasi-state, quasi-corporate interests. Johnstone, for example, mentions the fact-checking role of the US state-funded Atlantic council. We might talk more broadly of the ideological apparatus of the state, e.g. the so-called “foreign policy blob”.

Yet in other senses, the relative distance of the formal state matters. The censorship is wielded to defend specific, not general accumulation. Directly, the algorithm is used for the defense and advancement of particular companies, not capital overall. Indirectly, of course, it serves all these things, but in the most final and direct sense, shareholder value is king.

Sixthly, this censorship doesn't have clear, publicly known, targets or goals: When we think about censorship we often think of circulated lists of banned material or topics, yet algorithmic censorship is very different. Often, the companies hide the details of what the algorithm wants. This is for many reasons, one of them is so that people can't hit it with content perfectly designed to be on target, but not in line with the spirit of what they are looking for (c.f. Goodhart's law). In other cases, they might make aspects of the algorithm very clear- e.g. during the disastrous attempt to pivot us all to video. But, to a large degree, we are not meant to know the rules that prioritize content, even as these rules govern our conversation.

Also unclear- is the extent to which companies are motivated by direct financial interests, versus the extent to which companies are motivated by longer-term interests of cozying up to state and political actors.

Seventhly, algorithmic censorship is so pervasive that it makes imagining an alternative difficult: Algorithmic selection of the content we look at is so pervasive that it's difficult to even imagine an alternative. What would a reasonable, politically neutral principle for serving up content on platforms look like? Most people seem unaware of the issue.

Almost all criticism of the political role of media companies has focused on bans, rather than their algorithmic powers. Almost no one is demanding an unfiltered feed. As the saying goes, we don't see it like a fish doesn't see water.

A different feed is possible. We can't escape algorithms- any content serving formula is an algorithm. But it would be possible to create an algorithm that works only on a bare, content-neutral minimum. Such an algorithm would consider only factors like views since posting, number of positive interactions with the post, number of comments, and so on. If these factors are too slim, one could also throw in a past success rating- how often has content posted by this user been popular in the past? I concede that there will be great philosophical difficulties in separating out the content-based from the content-neutral, but some things are clearly more in one direction or the other.

Eighthly, this form of censorship challenges liberal categories around censorship: The most obvious challenge is to the public/private dichotomy which animates liberalism. As we alluded to above, in a formal sense, Twitter is a private space, not a public square. Thus according to classical liberalism, it's fine to censor. Yet If a huge chunk of the conversation is happening in one place, is it fair to treat that as private, in the same way, a saloon might be treated? What about Twitter's returns to scale and network effects?

However, the conceptual difficulties in the liberal framework are deeper than this. Even the core conceptual category of censorship discourse- the idea that some things are censored and others are not, breaks down in algorithm world.

Another category distinction in liberal thought that gets put under tension in this context is the distinction between regulating form and manner and regulating content. I hope to expand on this in a later essay.

We might also wonder how applicable the Millisian defense of free speech is. Whether it could be updated and whether it should be updated.

Finally, this censorship is sold to liberals as a weapon against ignorance and the right, but above all hurts those to the left of the current Overton window:

The right has its own communication channels. Right-wingers want right-wing content and will seek it out, so media companies give it to them. Despite disturbing moves like banning Trump from Twitter, there are limits to what media companies can do against the right. Even rightwing stuff that is heavily suppressed- pandemic denialism for example- is popular enough that it can't be squelched quietly. Thus the right is mostly unharmed, if anything they wear it as a badge of honor. To the extent it has any effect at all, it may have been to weaken some of their sillier elements- e.g. Alex Jones.

The actually-socialist, left of liberal, far left, however, is a vastly smaller formation in American life. It is not well-rooted in the culture. Thus algorithmic censorship is a far bigger threat against it.

Resistance

At this point in the anthropological inquiry, it would be traditional to try to flip this around and tell a story of how the plucky resistance is subverting the algorithm and power is never total. No doubt that's partly true and will be for the foreseeable future. However, machine learning algorithms are going to get better and better, and megacorporations are likely to get more deeply integrated into the state. My outlook is gloomy.

I don't have solutions. But I do have three thoughts on possible strategies:

* Clearly, It is necessary to create websites that are, as far as possible, outside this hegemony. Some kind of independent platform that catches on. Maybe one of the open-protocol social media sites people have been working on. But I want to give a caveat, I'd say that one of the biggest mistakes that could be made would be creating an

explicitly politically branded platform, tied to the left, right, libertarian, or even radical centrist. Lord forbid I never thought I'd find myself saying this, maybe we need to tie into that contentless resistance vibe of the 90's. Create something that is genuinely politically open in its amorphous opposition to "the man". I never in my life thought I'd find a practical use for this sort of Ad-Busters politics, but this might be it.

* We need to get clearer on exactly what is we're objecting to, and what our alternative is. One can't say that one objects to algorithms selecting to content one sees because every possible method of selecting content by computer is algorithmic. Instead, we need to start thinking about what we're willing for an algorithm to do, and what we're not willing. We need to start envisaging what content-neutral or content fair algorithms could look like. An algorithm that only considers metrics such as likes and views is one possibility, but there are many possibilities to explore. For example, giving every user the choice of which third-party designed algorithms they'd like to employ.

* A friend of mine has suggested that we should demand the creation of a national social media infrastructure owned by the state. This would by no means end the censorship, but it is easier to hold the state accountable for content choices than quasi-private monopolies that are thick as thieves with the state. A provocative suggestion well worth discussing. I guess that a state-run platform might be less indiscriminate and prolific in what it cracks down on, but when it did crackdown, it would crack down harder. Still, worth thinking about, though on balance I am opposed.

Yvne: The forgotten opposite of envy

Yvne (pronounced “Iv-Knee”) is the opposite of envy. Where envy is unhappiness that someone has done better than you, Yvne is the joy and satisfaction that comes from a sense of having done better than someone else. Yvne isn’t joy in having nice things, or having done great things, Yvne is joy in having nicer things and having done greater things.

If you’ve never heard this word before it’s because I had to invent it. Its closest pre-existing equivalent in English is probably the German loanword Schadenfreude, but this is not quite the same thing. Yvne does not necessarily require that a calamity befall the other person, or that their condition be miserable, simply that your success be in excess of their own.

It’s something of a mystery why I had to invent the word. After all, Yvne is no less common than envy -the rich indulge in it all the time—and it is no less objectionable than envy either. Indeed, I would argue it is worse. Envy pits your interests against those who are doing better than yourself and gives you a reason to drag them down. Meanwhile Yvne pits your interests against those who are doing worse than yourself—it gives you a selfish reason to prevent those weaker and more vulnerable than yourself improving their station—isn’t that more sinister than envy? At least envy is about punching up, yvne is all about punching down. Yet envy is denounced as one of the seven deadly sins, while people very rarely even talk about the concept captured by the word “Yvne”, despite its dangers and ubiquity.

You may have already guessed the reason I think people talk about envy all the time, but very rarely talk about yvne. The most powerful people in society have much to fear from envy, and so wish to condemn it. Meanwhile, these same powerful people enjoy yvne as a secret wellspring of pleasure. Back when religion held more importance, it was employed by the powerful to condemn envy. Now the task falls to economists. By

contrast, the opponents of Yvne have never enjoyed the same level of funding.

Postscript, why yvne cannot be identified with self-satisfaction or smugness:

I've had some people argue in response to this piece that yvne can be identified with either self-satisfaction or smugness. There are two main problems with this.

Firstly, I can experience yvne even while totally dissatisfied and dismissive of my own achievements. For example "I'm a total loser, but at least I'm doing better than my cousin."

Secondly, one can experience smugness without even thinking of others. I might feel smug and self satisfied upon completing a log-cabin in the woods, without even thinking about other people.

There are many concepts that are "almost" yvne, and overlap in many, even most cases (schadenfreude is another one, as is pride in general), but none are quite the same thing.

The paranoid style in petit-bourgeois politics

I've found that a lot of Qanon supporters and other rightwing conspiracy theorists are relatively financially successful and/or small business owners. The element of rightwing political affiliation is unsurprising on demographic grounds—these people are older, white, Republicans so why should anyone be surprised that they often own jet-ski dealerships and diners? But what about the conspiratorial element of these worldviews? Is there a way in which the lifestyle of a small business owner might lead to conspiracism?

It makes a lot of sense when you think about it. Put yourself in the position of a 60 year old, white property investor/car dealership owner. you're doing pretty well in the financial aspects of the game of life, but like a lot of people who are doing fairly well (especially older people who are doing well), you feel like you haven't gotten all that you deserve dammit!

But you're in a double bind. You can't question the rules of the game wholesale, because you want to believe that the game is in essence fair, insofar as it has put you ahead of many other people. That means any kind of left-wing critique of the justice of the social rules is out. You can't even really imply that the problem is ordinary breaches in the rules, like tax-evasion, because you've probably done those things yourself.

So if you can't question certain rules of the game, one strategy to explain the discrepancy between what you have and what you think you deserve to have is to imply that other people got ahead of you by breaking the rules- and not in the small ways you do. Elites must be totally disregarding the rules. Since normally breaking these rules is punished by society, there has to be some way a set of elites are circumventing them en masse. A secret, en masse violation of the rules sounds a lot like a conspiracy.

Maybe you think they're tied together by a blood that most people don't share

(anti-Semitic conspiracies), maybe you think they're tied together by an ideological project to circumvent and eliminate the economic rules altogether (anti-communist conspiracies), and maybe you think they're tied together by supernatural pretensions or even real supernatural powers (conspiracies about the occult and elite Satanism) Complicating the story I've told though is the reality that these people are not entirely wrong. The hyper-successful absolutely do breach the rules all the time, often through conspiracies, although this isn't the only, or main reason that these people are more successful than the conspiracy mongering petit-bourgeois.

This all might sound pretty niche, and I guess to a degree it is, but to the extent that both conspiracy mongering and the petit bourgeois are linchpins of contemporary MAGA-dom, understanding their overlap is important.

TLDR: The conspiratorial style in petit bourgeois politics arises from a fundamental conviction that they are hard done by—that you haven't got all you deserve. Naturally if you think you've been hard done by, you'll want to understand why. However, in explaining why you have been hard done by, you do not want to admit any fundamental problem in the underlying social order, because you have profited from that social order, so instead you must appeal to the idea that the social order is being circumvented. The most natural way to explain mass, gross, but hidden circumvention of the social order by everyone who is ahead of you and doesn't deserve to be is to posit a conspiracy.

Twitter is a reverse panopticon: The internal agent

Foucault in a frankly over cited discussion refers to Jeremy Bentham's panopticon—a hypothetical prison in which one never knows if one is being watched because there's one way glass everywhere. The possibility of being watched instills behavior similar to if one actually were being watched all the time. Eventually (and this is somewhat my own gloss) the hypothetical watcher becomes internalized within the “watched” person's own psyche, potentially unreal, but structuring their behavior and attitudes anyway. Twitter (and I suspect other mediums like Tik Tok & Youtube) is like this, except instead of fearing⁽¹⁾ that someone important might be watching we are hoping someone important might be watching.

This hope reshapes all online consciousness, eventually becoming a kind of internal, clout seeking voice. If the superego is an internal disciplinarian that arises from the external discipline of our parents, I call this figure the internal agent and picture him as the first result I found googling “sleazy Hollywood agent”.

The Rosetta stone to Twitter is that most of the accounts you see on your feed are hoping to be discovered. I held back from saying this for a time because I was afraid that it was really only a reflection of my own weaknesses—“most people aren't as narcissistic as you!” I thought. But I have slowly become convinced. Maybe not by the total number of people, but certainly by the total volume of tweets, far more than half of the people you see on Twitter are like the waiter in Hollywood who wants to tell you about his screenplay. I think we all know of this Rosetta stone at some level, but partly out of good manners (often an enemy of social criticism) we keep our awareness submerged. To reiterate, Twitter is an audition held in a panopticon and there's probably no one on the other side of the glass

This has a couple of effects. Firstly, people write in a way which is implicitly dissatisfied with its real audience, and aimed at an audience cooler and more popular than their

actual readers. There's something eerie about the fact that—at least to a degree- they are not talking to you.

Secondly, people have sensibly realized that because it's a very big audition they've got to stand out. Just being very good and incisive as a writer is not a great strategy for standing out. Too many people are playing that game—worse, too many people playing that game who also have something else to offer—e.g. good looks or a compelling life story. So if you want to win you've got to find a niche and that niche probably can't just be “is a good writer” unless perhaps you are very good indeed, but a lot more people think they are exceptional than are exceptional.

As a further consequence of point two, people are more hostile because it's an easy way to stand out.

Thirdly, interactions have a subtly strategic relationship quality. This is most obvious when people tweet things like “favorite this for a compliment” or “I reciprocate follows”, but those seem to me to be just the most obvious manifestations. This air of “you help me stand out and I'll help you stand out” is pervasive.

But fourthly, and above all, the joy of activity that is within itself -that does not point to some greater ambition—is tapered and in some cases eliminated.

You've probably already clocked that each of these four facets makes human interaction less authentic.

And all this for the sake of an internal mental observer that, as a general rule, doesn't reflect anything real.

(1) We are also fearing that someone might be watching and swoop down to cancel us,

but that's a discussion for another time. To briefly elaborate—not only are we seeking good publicity, we're also afraid of the bad. The end result is that we're doing all the work of celebrities with few of the rewards. Though not quite “none of the rewards” because there is a certain pleasure to the celebrity LARPing that Twitter and related offer.

Government Watch: An idea for an NGO

Government Watch is a hypothetical NGO that popped into my head today.

Government Watch sponsors a large jury (100+ members), picked in a way as close to sortition from the whole population as possible. The jury discusses matters among itself and seeks testimony from experts and affected people on whatever it chooses to pursue. It monitors government business and the issues of the day. Ideally, the jury is paid an honorarium to make taking some time off work possible. Possibly multiple juries would run at once. Thus far, Government Watch is similar to several experimental “citizens juries” that have been set up. However, its function is different.

The jury’s role is to monitor government action and decisions and, where appropriate, for want of a better term “call bullshit”. “Calling bullshit” takes a two-thirds majority vote. The two-thirds majority is necessary because if the margin were slimmer, it might just reflect the random composition of that particular jury.

“Calling bullshit” is the jury’s way of saying that the government has taken a decision that there is no chance that the population as a whole would support if they understood the issue. Basically, it’s a way of saying “look here, the government has done something awful”. The jury (hopefully) won’t call bullshit on decisions they think are merely bad rather they will call bullshit only on decisions they think are indefensible. The kind of decisions that obviously only serve narrow sectional and partisan interest.

A good example of a decision that I hope a Government Watch jury would call bullshit on is the decision to fold, time and again, to the tax complexity lobby. This is a real political grouping in America that lobbies to make income tax returns more painful.

The hope is that if you set this thing up, and did enough publicity for it, the fact that a Government Watch jury had called bullshit on something would, in and of itself, be

news, and that would be a punishment, however slight, for the government pursuing policies on behalf of obvious special interests. It would be essential for the functioning of government watch that it be non-partisan, that its juries be picked as close to randomly from the population as possible, and that there not even be the appearance of anyone exercising undue influence over the juries.

The jury might choose to go a bit broader than just targeting special interests, and go after especially cruel, inconsistent, authoritarian or wasteful acts it considered indefensible. However the two thirds majority requirement, and the representation of the whole population, means that we can be pretty confident that anything the jury called bullshit on had serious problems, at the very least in how the government had explained its actions.

It's not a cure-all. Most bad governance is better disguised than the tax-complexity lobby. However, I think there's a sizable and important niche here.

Notes on the Tyranny of Meritocracy by Michael Sandel

Part 1.

- It's obvious once it's pointed out, but it really is amazing how few people realise "anyone can succeed in America" is a big fuck you to everyone who hasn't succeeded in America. Some of those people vote!
- Sandel has gone through Trump's speeches and noticed that there are few, if any, sentiments along the line of "anyone can succeed in America " or "America is a land of opportunity" whereas there are hundreds of such statements in the words of his predecessors. He thinks this is not a coincidence, and I am inclined to agree.
- Sandel is right that a very interesting feature of political rhetoric is that it makes statements about the essence of how things are as a call for change—paradoxically. He gives the example of Obama saying words to the effect of "America is a land of equal opportunity, therefore we need to give every child a fair child a fair chance at an education". When you step back for a moment the two halves of this statement are completely contradictory! Yet in political rhetoric nothing could be more natural than to merge them. This trope of presenting a change as really a reflection of an underlying pre-existing essence is very old. C.f. Spartan political debates where people would interminably argue that their reforms were really restoring things to their constitutional essence.
- There are three political strategies for selling ideas and ideologies to people unhappy with their place in an unequal system, viz: A) The treasure your wins strategy—tell people that they should think instead about being happy to have beaten out those they lapped in the race of life, and get them on the side of maintaining inequality that way. B) The reroll the dice strategy. Say "Yes, you're right, there are some injustices in the system, so let's rework the rules to be a little fairer and then you, or at least your children, can reroll the dice—you might win this time! C) The gap reduction strategy reduces the size of the gap between winners and losers.

- At first glimpse it might look like A=Conservatism, B=Liberalism, C=Leftism. But it's not quite that simple—at least not all the time. Conservatives often offer a chance at rerolling the dice—“you'll be able to compete as a small business owner once big government corporatism is gone”. Liberals sometimes go for a `treasure your wins strategy—“those rubes want to devalue your hardwon education—don't let them, experts like you deserve to run the country”. Leftists mostly focus on C—gap reduction, but do talk about B—rerolling the dice—sometimes.
- Although many people have heard of the Great Gatsby curve by now, it's always worth a reminder that the dilemma between “equality of outcome” and “equality of opportunity” is fake because the two are deeply correlated.
- I've been thinking a bit about decadence lately—especially thoughts occasioned by Ibn Khaldun, Peter Turchin and my good friend Kieran Latty. I think one window into understanding what most people get wrong about decadence is the concept of “luxury”. People rightly associate luxury with decadence but for the wrong reasons. The reason luxury is associated with decadence is not because it's decadent to have nice things—at least in any meaningful conception of decadence. No, the defining feature of luxury is not having pretty or scrumptious or fragrant things—it's conspicuous consumption—trying to outshine your neighbors through your purchases. Decadence is a state of affairs wherein people -elites to be specific—view their primary goal as competing within society, rather than trying to advance society. I think this is Ibn Khaldun's sense of decadence—and he is right that it does destroy kingdoms and empires. This is also Peter Turchin's understanding of decline, and he is likely right that it is caused, in part, by an overproduction of elites.
- History isn't going to rap you on the knuckles because people are having a bit too much gay sex or men have long hair now, history is going to rap you on the knuckles if people aren't committed to larger projects than themselves. One of the main symptoms of that is luxury—people start buying nice clothes not because they are nice, but because they are nicer than yours. One of the great

disservices the right has done us is tainting the concept of decadence—a very important concept—and making it merely a vehicle of bigotry and trad aesthetics.

- One way to understand this is in terms of a Marxist theory of the state re: America right now yes I promise this will sweep back round to Sandel. So the US is doing very poorly—low growth rates, much unrest etc and a lot of it is clearly a result of bad governance. The Marxist theory of the state suggests that the state is the steering committee of the capitalists as a whole, and this is to my mind, largely true. Only that steering committee is meant to provide a synthesis of those interests. Obviously there will be conflicts between industries, but the state is meant to rise above that, at least “in the main”. However, the US state is failing to do that. It’s not an articulated whole of corporate interests—it’s a sack full of them – many contradictory—stuffed in greedily. That’s political decadence, and it is linked in manifold ways to individual decadence.
- Sandel’s point can be understood as—meritocracy is the ideology of decadent elites, so obsessed with their internecine competitions that they have forgotten their obligations to the common good and to the weak, so puffed up on their little games that they think fairness within their little club—and fairness in the rules of admission to their club—is the big game of governance.
- I want to end this section by reiterating that none of this means that either Sandel or I don’t think every child should have a chance to succeed. On the contrary, if you care about that, paradoxically, stepping back and focusing instead on making sure everyone has decent living conditions might be the best way to achieve your end, because the empirical evidence shows there is a strong correlation between low economic inequality and equality of opportunity. Meritocracy abstracted from this broader social project is—at best—about putting guardrails down to ensure that the children of quasi-elites have a fair shot at becoming full elites.

Part 2.

- A reader asked of the last post “what exactly do you and Sandel mean by this “meritocracy” you are critiquing? You’ve said you’re not opposed to the idea that the best person for the job should generally get that job, so in what sense are you not meritocrats?” That’s a great question. Sandel doesn’t quite spell out what he means by meritocracy, but I think there are a few separable but related theses, viz: A) Meritocracy as a theory of desert—people deserve remuneration corresponding to the degree of contribution their talent allows B) A theory of political priorities—the most important thing is to ensure “equality of opportunity”. Making sure that people at the bottom have a decent life is of secondary importance (if even that). C) A mode of rhetoric, focusing on equality of opportunity, the virtues of rule by the smart etc. D) A willingness to concentrate honor and dignity in hands of the “meritorious” e.g. talented. E) A theory of how political problems arise, viz not having the “best people” with the “smartest ideas” take care of them—instead of conflicts of value, practical interests and moral beliefs. Hence meritocrats find themselves committed to what I and Scott Alexander have previously called “Mistake theory” and contrasted with “conflict theory”.
- Sandel makes a big deal out of what he calls “smart language”, especially the language of “smart ideas” and “smart policies.” Such language is attractive during a period of political polarization where debates over what is right can seem so much more intractable than debates over what is clever, but the solution is a false one. Calling your policies the smart ones insults the intelligence of your interlocutor. What started as an attempt to avoid rancor feeds into it. This is a great point about mistake theories of political conflict generally—in trying to avoid conflict they risk inflaming it by implying that one’s opponents just aren’t as clever!
- Through analysis of rhetoric Sandel argues that the Obama era was a great time for mistake theory (although he doesn’t use the term of course). However, Sandel

also makes the point that if mistake theory was dominant during the Obama era as a mode of rhetoric and form of ideology that doesn't mean it was any truer as an analysis of the political conditions. Rather, constantly talking about "solving problems" "commonsense solutions" "smart solutions" etc. may have predominated precisely because politics was a morass of endless bickering at the time. The mistake theory rhetoric was wishful thinking and/or a futile effort at peace making.

- One of Sandel's more insightful points is: the obvious unfairness of unmeritocratic societies at least gave a "handle" that social critics and popular movements could grasp onto in an attempt to fight for a better world and better conditions for the lower classes. Critics could say "you have more than me and that's arbitrary, give me more" and there could be a conflict over that demand. Meritocratic societies on the other hand are "frictionless" in a way which doesn't dissipate popular anger, but instead leaves it inchoate—and potentially more destructive. Just because the apparent "fairness" of the system compared with overt aristocracy (not that the system really is fair, even on its own terms) makes articulating anger harder, doesn't make the anger go away.
- One of the most important points of the book is that people are, now more than ever before, concerned as much with the distribution of honor as with resources. This is something that, sad to say, I think the left has often gotten wrong historically. People are as worried about getting their lives to fit a meaningful narrative in which they matter as they are about making sure they'll always have food to put on the table. This may seem like a very obvious point—and surely on some level we all know it—yet I must admit that I've often failed to fully get it. Articulating a form of historical materialism which is fully alive to this need is important.
- Sandel discusses the history of meritocracy at length. Two things that stood out especially to me in his discussion. 1. Meritocracy may have been a cold war innovation—a desperate society turning away from entrenched privilege to ensure the best and the brightest would be in place to fight communism. 2. A

president of Harvard who was one of the original champions of meritocracy, and insisted it was distinct from equality of outcome, nonetheless couldn't prevent massive favoritism towards legacy admissions precisely because the alumni were rich and powerful enough to get their own way. This is a superb example of how massive inequality of outcome will tend to eat away at equality of opportunity—whatever noble intentions to keep them separate.

- Sandel reminds us of Erica Scharrer's fascinating studies of inept men in sitcoms. Over time there has been a tendency to portray men as more foolish—bumbling etc. in sitcoms, and women as competent and holding the family together. However the impact has fallen unevenly, with working class fathers much more likely to be portrayed as incompetent and useless and the tendency has been increasing overtime. This, suggests Sandel in conjunction with a variety of other evidence, is part of a pervasive cultural denigration of all those without a college degree, but perhaps especially men without a college degree. Sandel even suggests that the massive spike in deaths of despair among people—especially men—without a college degree may be linked to this general cultural denigration.
- This ties into something I've been thinking about for a long time. One argument we often see on the left is that it's okay to casually tease men, engage in joking (and even not so joking) misandry etc. because men aren't an oppressed group. Now there's a lot of truth to this, but it neglects another truth—when you attack a group, the brunt of that attack will fall on the weakest and most marginalized members of that group. Rich and powerful men will laugh off criticism of men in general. The people who get hit will be poorer, lower status men. The same is true of attempts to “discipline” the bad behavior of men or other dominant groups. The more powerful members of the group will often evade discipline, and it will instead fall upon the less powerful members—poorer, more likely to be disabled etc. A more sophisticated approach to social structure is required!
- An extreme example of how “castigation of the privileged” can harm the vulnerable—those people who thoughtlessly say things like “I can't understand

white men who still end up homeless, you had everything going for you and you still failed”. A disgraceful sentiment.

- Sandel argues there has been too much emphasis placed on distributive justice, when really we should be equally interested in contributive justice. Everyone wants to feel like they are making a contribution to society. At least at the margin, people’s most pressing unfulfilled desires are often not about consumption, but about feeling like they are making a valuable contribution to society. A lot of why inequality stings is not because it means we can consume less, it’s because society is quantifiably scorning our contribution. An ethic of competitive meritocracy doesn’t make the losers feel like their contribution is very significant. Politicians, economists and political philosophers alike have been guilty of making people’s identities as consumers primary over their identity as producers.
- When I think about my own greatest fantasies—to be an acclaimed writer, singer or philosopher, to be a hero, it’s notable that they are all fantasies not of taking from society but of giving to society and of being recognised for that contribution. I don’t think I am unusual in this regard.
- It’s worth noting that there are resources within my broad intellectual tradition—Marxism—for recognising and addressing exactly this point. The idea of the producer alienated from his product in any number of ways is just as fundamental to Marxism as the idea of material scarcity.
- Contributive justice might be all very good and well as a goal now, but let’s say that AI gets better and better and consequently the value of many people’s labour falls. How can we aspire to give everyone contributive justice under those conditions? Sandel doesn’t grapple with this problem, but I think it’s an interesting one. Let’s say that the transhuman solution of “upgrading” everybody so that they can make a material contribution isn’t viable—at least for a time.
- I think under these conditions the best we could probably do would be to encourage people to see themselves as contributing through actions and ways of being that are inherently meaningful. Joy, friendship, self-discovery, making art.

To shift from contribution through the production of extrinsically valuable goods to contribution through the “”production”” of intrinsically valuable goods. Another good that people can provide that doesn’t necessarily depend on skillfulness is giving their own preferences about what is ultimately, non-instrumentally good in democratic deliberation. I’m not saying it will be easy, but I think there could be a path to give people a sense of making a meaningful contribution even in a post-scarcity society.

- One thought of Sandel’s that will stick with me—the writing is on the wall for neoliberalism as currently understood. Even its most ardent supporters should be able to recognise this by now. The question then is not will the present “mode” of capitalism fall apart, but what will replace it? Authoritarian centralism? Quasi-feudalism? A replay of the post-war years with renewed unions? War and barbarism? Literal fascism? Social democracy? Socialism? I don’t know if the future is open, but it is certainly unknown. All we know is that the present won’t last. Understanding that the tower is going to fall, we just don’t know which way yet, is an important shift in perspective.

Why don't people often try to earnestly persuade others of their political views on the internet?

Scott Alexander has an article, entitled with Guided by the Beauty of our Weapons in which he makes two observations which have always stuck with me, even when I couldn't remember the name of that article:

A) Startlingly few people online make a bona fide attempt at making a case for their political beliefs that is likely to persuade anyone who doesn't already agree with them.

B) When you make such a case, in the right context, in the right way, people often react very positively. I can back this one up myself from personal experience.

But this creates a mystery- if persuasive material can work very well, but is rare then why are all these damn \$50 bills lying on the sidewalk?

Some definitions. By persuasive material I mean something relatively simple:

Online material which is genuinely intended for an audience that does not already share your basic beliefs on a subject, meant to bring them closer to your beliefs. This simple definition implies some fairly obvious riders, for example, persuasive material shouldn't be abrasive, let alone abusive, to people who don't already share your beliefs.

There are only two types of adjacent material that I want to take pains to exclude.

Firstly, political education, that is material intended for people who already share your basic framework for viewing the world, elaborating on some aspect of that framework. Persuasive material must be intended for those across a reasonably big political divide, political education for other believers doesn't count.

Secondly, “fact checking” and polemical “corrections”. These can be very useful, and, like political education, may sometimes incidentally persuade, but they are not what I have in mind.

Plainly, whether something is persuasive material, political education or polemical fact checking will often be a matter of degree, and many pieces will do more than one, but I nonetheless find these distinctions useful.

To be clear, I am not dismissing the importance of material which is, on this definition, unpersuasive. I am merely bemused as to why persuasive material isn't a bigger portion of what is written- why is it not 20% instead of 3%?

Here's a way of setting up the paradox. Consider the following premises:

1. The two most fundamental ways to achieve political goals (during peacetime) are organizing and persuading.
2. A lot of people online are interested in politics and have political goals.
3. People aren't stupid. They're strategic actors when it comes to what they want, and it's unusual for large groups of people to leave \$50 dollar bills lying on the ground.
4. The vast majority of political content we see online isn't primarily intended to persuade those not already convinced, nor would it be fit for that purpose.

There is, if not an outright contradiction, at least an apparent tension between these four premises. In particular, it's hard to see how, if 1-3 are true, 4 can be true.

Or to put it in a personal way. I am not an especially charming person, perhaps I am a little more charismatic than the average person, but I am nothing special. Despite that, I have, simply by making a consistent effort, shifted the political beliefs of several score people quite substantially. That tally only includes people I've met personally. Who knows how many people I have influenced but never met. So why aren't more people doing this sort of thing, especially online?

In this essay, I'm going to give some explanations. I believe every explanation here holds an element of truth, but I will grade them out of 10 on how much truth I think they contain.

Explanations that challenge the premises of the question

Persuasion doesn't work theory

One very simple explanation for why people aren't trying to be persuasive is that persuasion doesn't work and people know this. I don't put a lot of credence in this theory. I wasn't born a socialist, nor was I raised one. Other people weren't born conservatives. Nor did I spontaneously recreate socialism on my own, people explained it to me. Thus, at some point, there has to be some persuading going on. Still, there is some truth to this theory. Persuading people of things is difficult, and even more than difficult, it is frustrating (especially if you approach it the wrong way). I rate this theory 5/10. There's a dollop of partial truth here.

As an aside, I will add that, perhaps because we're out of practice, when people do attempt persuasion they often do it in a hacky way. It carries the tone of "I am here to educate you in a whiny voice, I am probably wearing a lanyard and will probably say "well actually". This could be leading to a feedback effect of persuasion continuing not to work so well because we don't have many good models for it.

Persuasion doesn't look like you think it does theory

A more sophisticated version of the previous theory is as follows- people are trying to be persuasive and persuasion does work. However, contrary to what you might expect, the best form of persuasion is bombastic statements that look, on the surface, like they are just preaching to the choir. Persuasion, then, is already all around us.

We might take a quasi-Calvinist view on this. Perhaps all the people who are going to become conservatives or socialists, or liberals or whatever already have it, deep in their blood. That is to say they are already the elect of these ideologies. Thus the right way to persuade is not so much to try and cajole liberalism, socialism or conservatism into someone, but to merely uncover what was already there. It just turns out that clear, eloquent, and above all forceful statements of opinion are the best way to do this.

I don't dismiss this entirely. One of the things that dealing with Trotskyists at university impressed on me is the power that comes from having a very clear, explicit line. Yet at the same time I am dubious, because those same Trotskyists, who seemed to have a definite ceiling on their success, impressed on me that clarity and passion alone will only get you so far. Also, even if clear, fiery statements are generally the best way to pursue persuasion, surely there should still be some more room for trying other things? My own experience of a more measured approach suggests there are at least some out there who will receive this better. Yet the landscape seems to grow close to a monoculture of stark and fiery statements. I rate this theory 7/10- interesting, provocative, but at best partial.

Explanations grounded in the psychology of individuals

Signaling theory

Or what about signaling theory? According to signaling theory, people aren't trying to persuade others because they aren't really interested in doing politics or achieving political goals. Rather they are only interested in signaling that they have the right politics. Persuasion is a bad way to do this, it requires an openness to dialogue. That creates a sense of ambiguity and complexity which is bad for signaling affiliation.

Now this is a very good theory, and I think it is a lot of the explanation, but there are a lot of people in the world who I truly think would, in a non-showy way, die for what they profess. There are others who I am confident would make smaller, but still meaningful sacrifices. Yet strangely, even these people often don't seem to be making a concerted effort to persuade others. Overall I rate this theory 8/10.

Vicarious expression theory

Similar to the signaling theory we have the vicarious expression theory. People are doing politics online to vent frustration, not to achieve political goals. Persuasion is bad for venting, thus it doesn't happen much. The reason they are not trying to attempt political goals is open to further debate- maybe they don't really believe in their stated goals, and just want a rival sports team to yell at, or maybe they do believe in their goals, but view their internet time as "downtime" when they're not really trying to accomplish anything except venting. I rate this theory about a 7/10.

Incapability theory

According to incapability theory, people don't attempt to persuade others very often because they lack the capacity. Exactly what this incapacity consists in is open to

debate, maybe they lack the confidence to attempt it, the patience to keep trying or maybe they lack the knowhow of persuasion?

Now there's probably some truth to this as an explanation, but we must remember that stuff we see on the internet is generally written by unusually erudite and confident people relative to the internet baseline. Thus even if a lot of people, or a majority, are incapable of being persuasive, there should still be more than enough who are capable that we should see their work everywhere, but we don't. Also, since when has lacking skill or confidence been a major barrier to anything on the internet? I rate this theory 5/10.

Explanations grounded in information ecology

Priced out of the attention economy theory

One very simple explanation of why we don't see a lot of content intended to persuade on the internet is that pieces that try to persuade take a lot of time and energy to process and so are too pricy to be widely bought and sold in a tight attention economy. I call this the priced-out of the attention economy theory. One plausible objection to this theory is that there has always been a niche market for long-form pieces on the internet-longish youtube videos, long-form articles, etc. If this theory is true, why aren't more of these aimed at persuasion?

One line of reply is that quite a few of them are. Think of Natalie Wynn's long-form video's for example- these seem like they belong to the handful of cultural productions which make an earnest attempt (often a very good one!) to persuade. So the priced-out of the attention economy theory has two virtues- it can explain why there are few attempts at persuasion, and it can predict where those few attempts that do exist will be found- with the long-form content that somehow survives despite the internet's preference against it.

A more fundamental objection to the priced out theory is that even on Twitter it is possible to say something that aims to be persuasive in 240 characters. It's hard, it doesn't come naturally to the medium, but it is possible. There is such a thing as persuasion in brief, thus a bias towards brief media can't entirely explain the parlous state of persuasion.

I rate this theory 8/10. I think it holds a very substantial portion of the truth, but not all of it.

Persuasion doesn't work online theory

What if there is some particular aspect of the online experience that makes it difficult to persuade people online, that doesn't apply to other contexts? Perhaps, for example, the problem is that when people have a sense of tension or uncertainty that often arises in the context of persuasion, they have the option of clicking out immediately.

I'm not quite sure what to make of this one, it doesn't seem to me to explain the few times people have tried persuasion and it has worked quite well. Still, I think it probably has an important component of truth.

I rate this theory 7/10.

Natural selection theory

Now we come to perhaps my preferred explanation for the lack of attempted persuasion- how stuff that gets circulated, is chosen, and the "natural selection" of content. The content that we see a lot of, content that gets shared, upvoted, and favorited, is chosen on the degree to which people strongly approve of it. But, by definition, material that is meant to be persuasive isn't often strongly approved of, for one thing, it's primarily targeted at people who are not already believers, and thus unless it is wildly and instantaneously successful, it is unlikely to leave its target readers

enthusiastically agreeing with it. Thus we can say of its audience- even if they upvoted it, they are unlikely to share it, and even if they share it, they are unlikely to share it multiple times. This means that we are unlikely to see much of the persuasive material which is created, and in response to this market signal, less of it is produced.

This is my personal favorite theory, I rate it 9/10.

Closing advice- swim against the current

My personal advice would be that if you really care about the ideas you believe in, you should try to swim against the current. Despite the odds, I've found it does work sometimes, and when it does, it's a lot more meaningful than screaming into the void.

The mysteries of welfare economics

There is a whole discipline that profoundly shapes government decisions everywhere. Even some very educated people know nothing about it. It is laden with philosophically and politically interesting and debatable premises. Despite this, some relatively modest technical barriers mean the public scarcely ever talks about it. It's called welfare economics, and in respect of how it influences our life, it may be the most important subfield in the social sciences. It's obscure even in the academic world. Even among working academic economists, it's sometimes considered a bit of a backwater.

The antinomy of welfare economics

Consider two plausible claims:

- The government shouldn't spend more money on services or infrastructure for one person, just because they are richer than another person. A government might justly choose to provide additional expenditure on essential services for the poor (e.g. to deal with disadvantages) but they shouldn't do the opposite. To spend more on health, infrastructure, education or whatever else for the rich rather than the poor would be an affront to both democracy and fairness.
- It is generally preferable to redistribute through direct payments, rather than the provision of services.

These two premises are actually contradictory (for reasons I will explain in the coming paragraphs), yet they are both eminently supportable. Hence they are an antinomy. I call this the antinomy of cost-benefit analysis. Let me explain why they are contradictory and why I call the contradiction by that name.

There's a procedure used by almost all governments called cost-benefit analysis. Consider some project called P. Let's say P is building a new bridge. Version one of the project involves building the bridge at location Oneish. Version two of the project involves building the bridge at location Twoish.

Now we work out how much it costs to build the bridge. Let's say that, in this particular case, the cost constraint is irrelevant except that we can only build one bridge- either project one or project two- that is to say either at location Oneish, or location Twoish.

We approach the problem by working out how much people would be willing to pay for project one, and if there is anyone who would be willing to pay to prevent project one and then doing the following sum:

(Total amount all people would be willing to pay to have the bridge built at Oneish
MINUS total amount people would be willing to pay to not have the bridge built at Oneish)

Then do the same sum for project two, and determine which sum is greater. That is to say determine whether net willingness to pay is higher for project one or project two.

Now consider how this might work out in practice. To simplify things we'll assume a case in which no one would pay to stop the bridge from being built in either location. In the town of Oneish, there are three hundred people who are all willing to pay 500 each to have the bridge built there. This may not sound like much, but it is a lot of money for the impoverished residents of Oneish. Meanwhile, at the hamlet of Twoish, there is a small community of multi-millionaires- 20 people- who would each be willing to pay 10,000 dollars to have the bridge built there. This might sound like a lot of money, but to them it is a trifle, representing the slightly greater convenience that will accrue to them if the bridge is built in the hamlet of Twoish.

$(10,000 \times 20 = 200,000) > (300 \times 500 = 150,000)$. So the maths says we should go ahead with project two. Hopefully, this illustrates how cost-benefit analysis favors rich people because rich people, having more money, are often willing to pay more money for projects than poor people.

Now there is a way to redress this. Rather than looking at how much money people are willing to pay, we can look at how much money people are willing to pay but weighted according to their income. The simplest weighting, though not necessarily the justest, is proportional weighting. To do a proportional weighting we look at how much each person is willing to pay as a proportion of their income. Simple, right?

Here's the rub, at least according to some economists. If we adopt weighted cost-benefit analysis the tax rate on the rich is effectively raised. This will, presumably, disincentivize labor shrinking the economy. The argument goes that if the government isn't more likely to build bridges in rich neighborhoods, the people, being rational agents, will notice this. They will therefore find the idea of becoming rich less attractive because being rich will come with fewer extra goodies. Therefore, they will be less likely to work that extra hour or go to extra trouble to get a qualification. For these economists, using weighted cost-benefit analysis is much like raising the highest marginal tax bracket in terms of its effects on labor supply and the economy.

If you're anything like me you probably reply "well, so what, we need to raise the tax rate on the rich anyway". Fair enough, but the problem is that, for various reasons, many economists think that while redistribution is well and good, they would prefer to do it through taxes and transfers, rather than in kind. There are some mathematical results that seem to suggest that giving people cash, rather than giving them goods and services, will be a more efficient form of redistribution. Thus, these economists propose, we shouldn't try to redistribute wealth through goods like bridges, roads, parks, hospitals and schools. Instead, we should use standard, unweighted cost-benefit analysis

and hence give the lion's share of these things to the rich, but generously redistribute through taxation.

One advantage of this approach, I will concede, is that it allows us to entirely separate out two questions: A) The question of how redistributive the government should be B) The optimal level of public goods provision. In a manner of speaking, it delinks questions of the size of government from questions about distributive justice. Yew Kwang Ng makes this point.

Now I have numerous problems with this argument for giving the rich more stuff, but I'm going to focus on two here. The first problem is an empirical problem, I'm not convinced that the rich experience, say, having extra parks in the same way that they would experience having extra money. For example, I'm not convinced people make the calculation that they're going to become an anaesthesiologist because there are more parks for rich people. At the very least, I anticipate that, as a matter of behavioral economics, the number of parks is less of an incentive on the rich than whatever the top marginal tax rate is.

The second problem is a matter of political philosophy. Unweighted cost-benefit analysis is effectively a form of voting in which the number of votes you have is determined by how much income you have. In other words, it is a form of direct plutocracy and, if you're anything like me, you probably think that plutocracy, whether direct or representative, is an unconscionable way to run a government.

But I'll freely admit, I don't have all the answers here. That's why I call this problem an antinomy. There are both really good reasons to think that redistribution should be done through transfers rather than in kind, and also really good reasons to think that building more parks for rich people than poor people is unconscionable. At the VERY least, building more hospitals and schools for rich people than poor people certainly seems terrible.

Although intellectually I am uncertain, my heart makes me fall on the side that says unweighted cost-benefit analysis cannot be allowed to stand. I am very wary of rationalizations for giving more things to rich people. I think such rationalizations will tend to get more support than they deserve anyway- whatever their intellectual merits. The rich will always have their ideologists to speak out in favor of them.

Here's what I think is going on here with this antinomy. There's a tension between capitalism and democracy. We talk about liberal democracies or capitalist democracies like they were the most natural thing in the world, but democracy and capitalism are two two very different organising principles that we have allocated different roles to. The market can be seen as a social planner in its own right. Assume that marginal utility in income is decreasing, so that your 100,000th dollar matters one quarter as much as your 50,000th dollar. As Brad de Long points out:

“The market system's social welfare function gives each individual a weight inversely proportional to his or her marginal utility of wealth.”

Insomuch as, from the point of view of the market, your 100,000th dollar is worth as much as my 50,000th dollar, even though my 50,000th dollar matters four times as much to me, your consumption preferences are weighted four times more heavily than mine in deciding what gets produced and who consumes it. Capitalism is an oligarchical system for making social decisions about production and consumption.

Public goods like parks etc. are things which it is not, generally efficient for individuals to purchase for themselves. Thus the government makes purchases on our behalf. But this brings the normally concealed tension between democracy and capitalism into the open because there's a border dispute between capitalism and democracy as social ordering principles. The purchase of public goods does not clearly belong to the oligarchical domain ordered by the market or democratic domain which is (in theory at

least) ordered by the electorate. We don't know whether buying parks or schools should be treated more like buying cars, except collectivized to avoid the free-rider problem, or whether buying parks or schools should be treated as something more like drafting criminal laws- something that shouldn't be influenced by money. That's why this antinomy arises- it reveals the seam between capitalism and democracy as social choice principles.

Making the perfect the enemy of the good- Welfare economics and social welfare functions

As you can imagine, unweighted cost-benefit analysis is controversial, because it effectively implements a direct-oligarchic voting system, in which each person's power over social decisions is proportional to their income.

Yet nonetheless, despite a great deal of skepticism among economists and political philosophers, most governments around the world tend to use unweighted cost-benefit analysis, although the UK is a notable exception, using weighted CBA for many purposes.

The rationale for this? Well, there are numerous reasoned defenses, including the one I talked about above. But I don't think these intellectual defenses are what has ultimately been effective. I think what has been effective is the fact that weighting cost-benefit analysis by some factor of income would oblige you to take a stand on values. Unweighted cost-benefit analysis wins because it is the default you slide into if you don't set up a weighting system. Even though it is far from value-neutral.

The conversation tends to go like this:

Aria: I propose that we adopt a roughly utilitarian weighing system.

Bravo: Aha! You say we should aggregate desires or happiness. But any comparison between the desires or happiness of two people is purely a matter of value judgment. There is no non-normative way to aggregate desires.

Aria: Come now, common sense, empirical psychology, psychometrics, and results from welfare economics like Lerner's equal ignorance theorem give us more than enough basis for reasonable comparisons between persons. In principle really not much different to comparing the temperature of objects.

Bravo: [Walks off disgruntled and clearly unconvinced]

Carol: Your utilitarian weighting system doesn't pay special attention to the worst off in society. We need prioritarianism.

Dylan: Your utilitarian weighing system ignores that there are higher things than mere pleasure and desire satisfaction. We need to use the capabilities approach.

Erol: I fell on my head as a child and for some reason believe that a Rawlsian maximin social welfare function is appropriate.

A figure returns. It looks suspiciously like Bravo, but with a glued-on mustache.

Not Bravo: Oh well this is just all too difficult, we had best just stick with unweighted cost-benefit analysis so we can make fewer value judgments.

Everyone all at once shouts no, but they are also shouting at each other. The treasury department resolves to use unweighted cost-benefit analysis because it looks easier than getting involved with all this mess.

When I first started working on my Ph.D., my solution was simple. Argue that all the critics of unweighted cost-benefit analysis had an interest in strategically adopting utilitarianism, because, from the perspective of all the popular alternatives, e.g. prioritarianism, the capabilities approach, maxmin, utilitarianism will be an improvement over unweighted cost-benefit analysis. What I was suggesting was a kind of political maneuver within political philosophy itself- rally around utilitarianism. The logic was that of the Pareto improvement- utilitarianism is better for all political philosophies at play than unweighted cost-benefit analysis.

This kind of argument is no longer really central to my thesis, but I still think about this a lot, viz: A) The way incumbent models can win out for lack of a unified opposition B) The possibility of forming political coalitions within normative political thought.

Making the perfect the enemy of the good part II- Welfare economics and psychometrics

Think about the inadequacies of willingness to pay as a measure of welfare.

The most glaring is that it treats the desires of the rich as many-fold more intense than the desires of the poor, but there are other issues as well. There are many technical problems with measuring it- for example, how should we work out how much people are willing to pay for clean water?

The simplest method would be to ask them “how much extra tax would you be willing to pay a year to get clean water”, but this verbal approach is considered suspect by many economists. Alternatively, we could look at commercially available water filters, see how much they go for, and value a population of people getting clean water as equal to how much it would cost to get them filters. Alternatively, we could look at their lifetime expected additional healthcare costs if they don’t get clean water, and value it at that- how much they would be willing to pay to get healthcare to deal with the consequences

of not having clean water. We could also look at how much they're willing to pay for bottled water. These and numerous other models contend together in the air, but there's no clear way to choose between them. People's willingness to pay is often context-bound, unreliable, and wildly divergent depending on how you look at it.

What are some other ways you could measure the welfare impacts of economic policy? Well, a natural suggestion is that we measure happiness using the subfield of psychology known as psychometrics.. Alternatively, perhaps we could adopt a hybrid model, where we use willingness to pay, but adjust it on the basis of various results from the study of subjective wellbeing. For example, if we know that happiness increases as a log function on income, treat the value of an extra dollar to a person as proportional to log income. My willingness to pay a dollar (at 50,000 income) is thus treated as equivalent to your willingness to pay two dollars (at 100,000 income).

But, the objections begin:

Which kind of happiness? Pleasure? The satisfaction of desires? Eudaimonia?

How do we know that the happiness scale is relatively constant between groups? [Never mind that we know for a fact that the willingness to pay scale isn't constant.]

What about (insert technical measurement problem here) [Again, this seems unfair, because while there are unknown technical details regarding how accurate the measurement of happiness is, we know that willingness to pay is bad.]

And once again unweighted willingness to pay wins out, due to a disunited opposition and institutional inertia.

Two senses of caring

Life is more about doing than feeling

-My dad

This is another one of those posts which is going to be painfully obvious to a lot of readers, but which others might find helpful. I wanted to talk through something that I only consciously grasped in my early twenties. The difference between two senses of caring about something or someone. It's important to be clear on this because being woolly about it might give you an unwarranted cynical view of people- and of yourself.

Once, attending the speaker's corner in Hyde Park, I listened as a speaker proposed that we don't really care about what happens to the vast majority of people. We might learn that 20 have died horribly, and shudder for a moment, but we'll sleep fine just afterwards. I objected that if the angel of death appeared before you and offered to give a reprieve for those 20 people in exchange for 200 bucks, almost everyone would take the deal, indicating that we really do care.

The speaker's point was echoed centuries ago by Adam Smith, in a quote often repeated:

“Let us suppose that the great empire of China, with all its myriads of inhabitants, was suddenly swallowed up by an earthquake, and let us consider how a man of humanity in Europe, who had no sort of connection with that part of the world, would be affected upon receiving intelligence of this dreadful calamity. He would, I imagine, first of all, express very strongly his sorrow for the misfortune of that unhappy people, he would make many melancholy reflections upon the precariousness of human life, and the vanity of all the labours of man, which could thus be annihilated in a moment[...] And when all this fine philosophy was over, when all these humane sentiments had been once fairly expressed, he would pursue his business or his pleasure, take his repose or his diversion, with the same ease and tranquillity, as if no such accident had happened.

The most frivolous disaster which could befall himself would occasion a more real disturbance. If he was to lose his little finger to-morrow, he would not sleep to-night; but, provided he never saw them, he will snore with the most profound security over the ruin of a hundred millions of his brethren, and the destruction of that immense multitude seems plainly an object less interesting to him, than this paltry misfortune of his own.”

But here’s the part of the quote that is not normally reproduced, it makes my point:

“To prevent, therefore, this paltry misfortune to himself, would a man of humanity be willing to sacrifice the lives of a hundred millions of his brethren, provided he had never seen them? Human nature startles with horror at the thought, and the world, in its greatest depravity and corruption, never produced such a villain as could be capable of entertaining it. ”

So if we measure care in terms of feeling, we care more about our little finger than the whole population of China. If we measure care in terms of action- in terms of the tradeoffs we’re willing to make- we care more about China. We might call the first kind of care F-Care (feeling care) and the second kind A-Care (action care).

Like all tendencies to action, A-care is primarily defined in terms of what you are willing to forego to make the action happen- the costs and opportunity costs you are willing to incur. A-care then is fundamentally about sacrifice.

Sometimes A & F care actually go in opposite directions. To my shame I have inwardly rejoiced in the misfortunes of my enemies. Nonetheless, if you’d given me the option, I would have gotten them out of their predicament- even made big sacrifices to do so.

The most important sense of caring by far I would submit is A-Care. Jesus of Nazareth implicitly recognised this when he said:

“Greater love has no one than this: to lay down one’s life for one’s friends.”

If love/care is defined in terms of A-care, and sacrificing your life is generally the greatest sacrifice you can make, Jesus’s statement stands to reason.

Even though A-care is more important than F-care, A-care takes up a lot of our mental space. Those who seek a disciplined spiritual mind have often been greatly distressed by the possibility that their motives are not pure, and have sometimes interpreted, for example, their longing for glory incidental on a greater desire to serve goodness as proof that their motives are not pure. For example, in *Murder in the Cathedral* by T.S. Eliot, the soon to be martyred Archbishop of Canterbury laments:

The last temptation is the greatest treason: to do the right deed for the wrong reason.

Because the thought of being a famous martyr is very attractive to him- as one who has admired martyrs, his heart glows warmly at the thought of being a famous martyr. There can be no question that he would go ahead with this even if he would not be a famous martyr as a result, but he worries that his rejoicing in the possibility of a famous martyr invalidates his good intentions. At least if *The Quest for Authenticity: The Thought of Reb Simhah Bunim*, this sort of struggle to bring feeling and action in line was a key feature of Reb Simhah Bunim’s life as well.

This desire for “clean” feelings which match my moral values has been an ongoing struggle for me. I’ve felt my heart sore at the news of things that I would quite literally give my life to stop, and I’ve felt my heart plummet at receiving what is, in terms of my considered views, great news. This disconnect in judgement isn’t just related to direct outcomes, sometimes it can be about evaluations. For example, I’ve felt myself swell with overweening pride emotionally, feeling confidence in my own abilities- while nonetheless I still wouldn’t take a prideful bet. Presumably, at some point, this begins to

shade into the subtle distinctions between categories of beliefs that I previously covered here.

My personal advice, after years of wrestling with the contradictions between these two kinds of care, is to stop caring about F care, and only care about A care when you're evaluating yourself. Easier said than done, and I acknowledge the possibility that I'm wrong, but trying this has worked well for me- especially in light of my self-critical form of OCD which has made clarity very hard to come by.

I have my own half-baked linguistic proposal to clarify the distinction between F & A care in contexts where it is very important to be clear about what you mean. I propose that for A-care we revive the Latin word "necessitas" which according to Google translate means: "necessity, compulsion, urgency, requirement, tie, relationship". For F-care I propose we revive the Latin word "sollicitudo", which according to Google translate means "solicitude, care, concern, anxiety, worry, apprehension". I have found thinking about myself and my actions in a way which clearly distinguishes them is useful in trying to understand myself, my strengths and failings in a realistic way.

We are, on the whole, what we do, not what we feel. If there's one other takeaway here, it is that you are more fragmentary than you may realise. A person is more like an interlocking field of gears, or the parts of a chariot than they are like an animated marble statue.

Relative income effects can be a lot of different things, not just envy

Relative income effects are effects on your happiness based on how your income stands in relation to the income of other people. If I am sad not so much because I have don't have enough money, but because I have less money than Tom, I am suffering a negative relative income effect. Relative income effects may be more significant than absolute income effects in our society. See, for example: *Absolute Income, Relative Income, and Happiness* (2008) by Ball & Chernova.

There is a perception that relative income effects are petty, ignoble, or mere envy. Some economists have even argued it is ill-advised to consider them in the context of economic policy-making for that reason. I wanted to take on that argument here by imagining many different ways your relative income might plausibly affect your happiness. As we will see, the vast majority are not envious or ignoble, at least at first glance.

A) Alexandria is less happy due to relative income effects. Because she is poorer than most of her neighbors, the thinning of the bottom section of the market for many commodities has reduced the variety of goods available for her to purchase. A greater portion of products sold in her area (or perhaps even her country) is aimed at the top of the market. For example, she can't find many restaurants or clothes shops in her area that cater to her price bracket. She is particularly worried about gentrification. If her neighborhood becomes further gentrified, she may be pushed out of her home of many years, unable to afford the rent.

B) Jason is unhappy because his low relative income means that he is regularly disrespected for the "shabby" standards of his clothes, house, absence of bed-frame, etc.

C) Isabel is similar to Jason, but she doesn't care about being disrespected per se, she's concerned about how that disrespect will affect her goals- like finding a partner, securing a job, being taken seriously in civil society etc.

D) Ebony is upset by relative income effects because she believes that income is a form of social recognition. Her relatively low income means she is not being justly recognized by society. In particular, she works as a hospital cleaner and does a lot of unpaid reproductive labor at home as well. Ebony feels that the millions of people like her deserve more. What she perceives as unjust pay stings her sense of self-worth.

E) Xi is upset by the high relative income of others because he believes these incomes give the rich disproportionate power, undermining democratic decision-making. This undermines his confidence.

F) If Xi's hunch is right, differences in income may mean he will be negatively impacted by decision-making which favors the rich- even when it doesn't show up in changes to absolute income (e.g.- decisions about everything from health policy to public park placement which will not affect income, but which will favor the poor over the rich- a relative income effect on welfare).

G) Marco is upset by certain very high incomes because he feels they have not been justly earned. He feels that the real contribution of, say, advertising executives to society - which may even be negative at the margin- is not at all proportional to their large salaries.

H) Theodora feels that huge disparities in income, and a focus on the top end of the market, mean she doesn't see enough depictions of people like her in the media.

I) Bob is concerned that the low relative income of his ethnic group contributes to stereotyping and prejudice against them.

J) Jessica is concerned about her low relative income because it means that she can only afford to put her children in a residualised school. Since schooling is more stratified by income in more unequal societies, and more prestigious tertiary education institutions tend to select students from more prestigious secondary schools, this will put her children at a disadvantage. Note that this is strictly about relative income- it's about the quality of the schools relative to each other, not their absolute quality. (Morgan 2021- personal communication)

K) Jennifer is lonely because she can't afford to participate in the social activities of their peer group. When they go out drinking, her low relative income means she can't join them. (Dannaher 2021- personal communication)

L) Simae feels instinctively humiliated by seeing many people of a much higher status than him, through a kind of automatic evolutionary reflex, or perhaps a response developed in early childhood. Noble or not, he can't control this reflex.

M) Lisa is concerned about her low relative income because she believes that her unique entrepreneurial genius in starting a jet ski dealership in Florida has not been recognized.

Of these feelings, I think only Lisa's can clearly be equated with envy. Some seem like clearly reasonable concerns, at least as far as they go, others will have to be debated on their own merits. Marco's claims about the salaries of advertising executives, for example, can be debated, but to treat them as obviously malicious envy seems wrong.

Why we can't (usefully) dismiss concern about the income of others as envious

I. The Easterlin paradox

The Easterlin paradox is the observation that country level happiness does not seem to increase overtime with income- even though there are many reasons to think it should- for example, income gives us choices, security, additional consumption and so on. The paradox grows even weirder when we reflect that individual increases in income do raise happiness.

In 2008 a group published a paper entitled “Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles“. The paper argued quite persuasively in my opinion that this phenomena can be explained through the relative income effect- that is to say much of the happiness we derive from additional income is to do with changes in our status relative to others. Consequently, simply increasing income won't do much- indeed if it is accompanied by increases in inequality it may lower average happiness, even if it increases everyone's income.

There are subtleties in how relative incomes are thought to work(1), but on one plausible model, relative income effects indicate that less inequality will mean higher aggregate welfare. If so, the existence of relative income effects give a reason to favour higher taxation and redistribution.

II. The moral case for and against disregarding “envious” preferences

Here and there one encounters thinkers who argue that relative income effects represent envious preferences that should be disregarded because they are morally illicit and are tantamount to a sadistic desire to level down. Although this argument hasn't yet achieved great prominence, as the literature on relative income effects as an argument

for egalitarianism becomes larger, objecting to “envious” preferences will no doubt become more popular as an anti-egalitarian defense.

It is perhaps not clear that relative income effects do represent envy. They may, for example, reflect legitimate concerns about their economic security or the security of democracy in the face of rising inequality, or legitimate outrage over a genuinely unjust distribution.

I want to argue though that regardless of whether relative income effects truly do represent envy, attempts to blame people for their “envious” feelings about the incomes of others are futile. They are futile because very plausibly, the need to not to be lower status than others is deeply hardwired into us. An ethics that enjoins us to ignore these feelings is simply stuffing its fingers in its ears.

In making this argument I am mindful that we must steer between the Scylla of the naturalistic fallacy (“X is natural, therefore X is justified”) and the Charybdis of taking no account of the composition of human psychology in considering how we should order society. Our argument is not that society should indulge these drives just because they are natural, but rather that that, because they are hardwired, if they not indulged they will nonetheless always remain as unmet needs. Since unmet needs prevent flourishing not indulging them would be a major impediment to flourishing. This is true even if we would prefer on the whole that such drives or needs not exist.

Does our scheme justify monstrous behavior? I doubt it. Let’s take violence as an example. If our scheme justifies violence that’s a major problem- I don’t think it does justify violence. We need to distinguish between the tendencies that evolution has given us and the needs and drives that evolution has given us. It is important for human happiness to make peace with our drives and needs. It is not so important to make peace with our behavior tendencies. There is a great deal of evidence that humans have a tendency towards violence in many situations, but with the possible exception of a few

psychopaths, very few humans have a need to be violent- they may choose to be violent in more circumstances than would be wise, but being violent so does not meet important and non-instrumental desires or make them happier. On the other hand, the desire to not be at the bottom of the ladder reflects a drive or need- not fulfilling it will lead to suffering, or at least a reduction in happiness.

Thus, insofar as we care about human flourishing, we have a reason to meet this human need not to be far below others. If we call this “envy” intrinsically evil or try to ban gratifying it, we condemn people to a less rich and flourishing life and will not, in any case, change this aspect of human nature. That seems like a good reason not to ignore these preferences.

Footnote:

(1): Note that if relative income effects are purely about rank order reducing income inequality won't help at all, at least not if any income inequality remains whatsoever, because in the presence of any level of income inequality there will be ranks. I find the notion that it's purely about rank order implausible. Also note that if the rich derive as much happiness from their extra status as the myriad poor derive unhappiness from their lower status, reducing inequality will not increase overall happiness. Again, I find this implausible. I doubt the zero sum model of relative income effects- it seems to me probably true that at high levels of income inequality more utility is lost to relative income effects, while at lower levels, less utility is lost. However some authors endorse a zero sum model, so caveat emptor.

Chesterton's fence and thinking using sayings

“Common sense is a chaotic aggregate of disparate conceptions, and one can find there anything that one like.”

-Antonio Gramsci

I.The fence

Chesterton's fence is the principle that:

“Reforms should not be made until the reasoning behind the existing state of affairs is understood. ”

(Courtesy of Wikipedia)

It sounds so very reasonable- simply a more specific statement of “look before you leap”. Like most entrenched bits of wisdom it is not precisely wrong. It is perhaps most plausible to anyone who has started operating in a new context- for example a job- and thought they'd found numerous absurdities and inefficiencies, only to slowly realise that most of them made sense for reasons that were not immediately obvious.

Like other bits of entrenched wisdom, it is an ideological iceberg, encoding far more contestable assumptions than are initially obvious. For example:

Institutions have a reason, and are rarely, if ever, just spandrels.

We can feasibly discern these reasons, or, at any rate, if we can't discern them for a particular institution, we have no need to change the institution.

(Implicitly) That the kind of interests that institutions serve are the interests of society as a whole, and not merely particular interests of classes, or even specific individuals.

(Implicitly) Change is an exogenous factor that potentially threatens the well adapted system, and isn't just another well-adapted product of the well-adapted system.

N.B. The points labelled (implicitly) aren't strictly implied, but tend to come as a bundle, and be mutually supportive with the overall mindset.

Generally speaking the overall picture is one of adaptationism and functionalism. Adaptationism is a family of views in evolutionary biology that tends to see a broad range of traits as adaptations to their environment, as opposed to contingency, genetic drift, etc. Functionalism is the view in sociology that society is like an organism, and its various institutions like functional organs within it, each serving some purpose of the whole, and not merely the particular interests of classes or other groups.

There's another subtlety here. Societies like ours have, for a very long time, been changing quite regularly and rapidly compared to many other societies for hundreds of years. Presumably then our society has venerable mechanisms of change and reform. Actions on society are not exogenous to society and its institutions so the image of the agent as pouring chaos from outside on a carefully balanced equilibrium is misleading. To put it slightly differently, if the reasons for institutions can be submerged and not immediately obvious, yet still vital, why not extend the same respect to reformers, who are themselves an institution?

II. Alternatives

We might counter-pose Chesterton's fence with various other principles, e.g.:

Marx's dike: Differences in both power and interests between people, mean that we should expect existing institutions, and the purposes for which they exist, to disproportionately favour the interests of the powerful.

Weber's walkway: Massive differences between countries in things like legal and institutional structure, as well as social outcomes, indicate contingencies and path dependencies matter a great deal.

Of course one can always say of a proverb that it is simply a tool for thinking, that it was never meant to be slavishly applied, and of course this is true. It is also true that no two people will agree exactly on what sensibly using it as a tool for thinking, and what slavishly applying it is.

One option would be to think without using these sorts of generalised sayings. After all, they frequently contradict each. Here is a list of contradictory proverbs from Liz Pullen on Quora:

Look before you leap or All good things come to those who wait / Those who hesitate are lost.

If at first you don't succeed, try, try again. / Don't beat your head against a stone wall.
Never put off till tomorrow what you can do today. / Don't cross the bridge until you come to it.

Two heads are better than one. / Paddle your own canoe.

Haste makes waste. / Time waits for no man.

You're never too old to learn. / You can't teach an old dog new tricks.

A word to the wise is sufficient. / Talk is cheap.

It's better to be safe than sorry. / Nothing ventured, nothing gained.

Don't look a gift horse in the mouth. / Beware of Greeks bearing gifts.

Do unto others as you would have others do unto you. / Nice guys finish last.

Hitch your wagon to a star. / Don't bite off more than you can chew.

Many hands make light work. / Too many cooks spoil the broth.

Don't judge a book by its cover. / Clothes make the man.

The squeaking wheel gets the grease. / Silence is golden.

A stitch in time saves nine / If it ain't broke, don't fix it.

Fools rush in where angels fear to tread. / Nothing ventured, nothing gained.

Seek and ye shall find. / Curiosity killed the cat.

The best things in life are free. / There's no such thing as a free lunch.

Save for a rainy day. / Tomorrow will take care of itself.

Life is what we make it. / What will be, will be.

Opposite attracts. / Birds of the same feather flock together.

Faint heart never won fair lady. The meek shall inherit the earth.

With age comes wisdom. / Out of the mouths of babes come all wise sayings.

Two is company, three is a crowd. / The more, the merrier.

But whether or not ceasing to think with proverbs would be good or bad, it won't happen. So what is the most productive way to use them? Contradictory bits of wisdom are disturbing if we think of sayings as things which are meant to be right or wrong. It is best, I think, not to consider them as right or wrong but as having different weights. The weights we place on each proverb will vary, but the saying serves as a kind of marshall for that side of the argument, and as a reminder that we should consider what weight we put on each.

Chesterton's fence, Marx's dike and Weber's Walkway are three different ways of looking at institutions. No one quite believes in one entirely, but the strength we place on them will vary. These weights will change based on our experiences and predispositions. Jointly they define a kind of space of possible views, emphasising respectively wisdom, conflict and contingency for drivers of how things are.

Just don't go changing your entire political practice for a proverb like Chesterton's fence without carefully interrogating it. If nothing else, that would be quite ironic.

Musings on the self-organization of authoritarianism

I. Rule through inaction, tyranny through silence

For some people this essay is going to be bleeding obvious, but it took me a while to get this, so maybe it will help someone. I've come to realize that most pro-authoritarian action can be self-organised- it doesn't need intervention from actual authorities, although it may benefit from such intervention. People simply act on what they perceive as their joint interests with the powerful, in the hope of social, psychological or material rewards.

We're all aware on of this process on some level, but I think it is worth moving to the forefront of our mind. What I believe, but cannot prove, is that it may be the largest engine of entrenched authoritarianism.

Some examples:

1. A wealthy man buys a newspaper. He never says "don't run negative coverage on me" but hey, the journalists and editors aren't stupid! No one wants to be the person who wrote an article criticising the boss. Maybe the boss even wishes they were a little more critical of him, just to create the appearance of neutrality- but none of them dare. Just to be safe, it might be best for the jurnos to moderate the criticism of billionaires generally.

2. A new government makes things easier for a specific industry for the purpose of encouraging investment. The industry has never told them to do that, it never offered a deal whereby it would invest more in exchange for regulatory changes, but politicians -rightly or wrongly- imagine it to be true.

3. The right wing fan of a politician starts a fight with a counter-protester at a rally. Publicly the politician has said they don't want fighting, but their loyal supporter imagines this is not their real position, which they must conceal due to "political correctness" or somesuch.
4. Larry works quite a bit of unpaid overtime. He imagines this will impress his boss. His boss hasn't asked for it. Perhaps he even wishes he wouldn't, concerned about possible liability and OH&S concerns. Nonetheless Larry puts in more hours, imagining he will win greater approval.
5. This one is more speculative, but I can't help but think that many authoritarians are trying to please some internalised father figure they've created by expressing the right sentiments, partially obviating the need for actual authority figures to tell them to do things.
6. Jessica wants to be an entrepreneur. She rails against socialistic government and the man in Washington who takes from the sweat of the business owner's brow. She joins her local young Republican chapter. She imagines this is what entrepreneurs think, and how they act, thus she believes she is making herself more like her heroes.

So, because they imagine it will win them either emotional or tangible rewards, or both, people organise on behalf of the existing authorities. The authorities need not lift a finger for this to happen, in some cases they might even wish it wouldn't. No paper trail is required. It's a somewhat dark take on the Taoist idea of ruling through inaction.

II. Letting others take the initiative on your behalf keeps off the stench of failure

As a strategy for ruling class individuals, saying little that is clear but allowing people to meet what they believe to be your needs, particularly suits those power figures who rely

on the appearance of infallibility. Letting others organise around your desires without explicitly intervening allows you to simply disown those actions which would embarrass or fail you.

Sometimes action by the powerful is synthesised with self-organising authoritarianism through the use of oblique and deniable statements and requests. For example, Henry the II's infamous utterance:

““Will no one rid me of this turbulent priest?”

As long as I can remember I've been aware that kings play this sort of game, but what I want you to consider is the possibility that this sort of game isn't just a feudal curiosity, it's the very substance of any system of power.

III. It's the f*** panopticon again.**

At the risk of being the cliché of the half-educated social science student, allow me to compare this to Foucault's adaptation of Bentham's concept of the Panopticon as a metaphor for social discipline. In the Panopticon one disciplines oneself by not performing transgressive actions, because one never knows when someone might be watching- even though it is unlikely they are.

Whereas the Panopticon focuses on punishment, the phenomena we are interested in here is a bit broader, because reward plays a role as well, indeed it is the larger part. Also, the shaped behaviour is as much about the concrete interests of powerful people as norms as such- this isn't just about power in the spaces between people, but explicit, enumerable ruling class.

Of course all this could be wrong, and this brief excrusus into Foucault is probably ill-advised because it's really not my area.

IV. Self-rewarding self-organization and rational kingly bubbles

Indeed self-organised authoritarianism may actually end up being rewarded- even without the intervention of an authority. It may self-organise its own rewards. After all, rewarding those who have served the master well seems like something the master would approve of. This means that even if you know better, participation in it might still be rational, a sort of social version of the economic concept of a rational bubble.

Yearning

Oh death, where is the antidote for thy sting? Or: Prolegomena to a new philosophy of the Common Task

It's about 2012. A friend of mine, about 30 years old, has just died of sepsis. I loved him, and he has been annihilated. I'm standing talking with another friend of mine who was also close to the deceased. A thought occurs to me. "Do you think we'll ever be able to fix it?" "You mean feel better? That will come with time." "No, I mean bring him back from the dead with technology." My friend looked at me in puzzlement and sympathy, thought for a moment, and said "No, I don't think so."

In the past when loved ones had died I had imagined death as a vast granite barrier which my hands could make no mark on. But what if we could find a ram powerful enough that the wall of Hades couldn't prevail against it? The thought seemed stupid, yet the future is long and holds many technological wonders. How could I be so confident there was no hope? A hundred years ago an eccentric, perhaps insane, Russian philosopher named Nikolai Fyodorovich Fyodorov suggested—on the basis of scant to no evidence—that such a thing might be possible.

I want to emphasise that I am not suffering from psychosis, so I do not really hold that the idea I describe here is viable. Yet I can't help but play with it and ponder it. Didn't we get where we are in part through mad dreams? To cheat a little with metaphors, maybe you need a vantage point some distance from what is possible to see the full scope of possibility.

I have a fantasy. I mean this entirely seriously when I say that I think it is among the greatest fantasies ever conceived. There is little vanity here because it is not my fantasy alone. What if we could redeem all of history—I really mean all of it. Give every story a happy ending by bringing the dead back to life. Not just slow or stop the advance of death, but reclaim each territory it has seized from us, and so, at least in a sense, correct every injustice there ever was?

My fantasy is a very old fantasy. It is essentially the fantasy of universal salvation. I'm an atheist, but it is typically a religious fantasy. It receives expression in Mahayana Buddhism and scattered forms of Christianity and Islam. I would bet good money that someone in the Jewish tradition has articulated it, but I haven't found a reference yet. I'm sure it can be found in many other places besides. Apparently it's currently a hot topic in Christian theology (or at least the Protestant strand thereof). You can find a trace of it in the Bible:

“On this mountain, He will swallow up the shroud that enfolds all peoples, the sheet that covers all nations; He will swallow up death forever. The Lord GOD will wipe away the tears from every face and remove the disgrace of His people from the whole earth.”

Emphasis is mine.

Generally speaking, the vision has been a supernaturalist one. In the absence of the supernatural it seems likely that people dissolve at death, with no directions about how to put them back together again retained in some secret archive. At least if the ancient philosopher Epicurus is any guide, this is what naturalists have believed since there were naturalists. There is at least one exception though—one person who thought salvation might be achieved naturalistically. In the late 19th and early 20th centuries, Nikolai Fyodorovich Fyodorov articulated what he called the Common Task.

Nikolai argued that one of the greatest forms of alienation stunting human potential is that of the living from the dead. The division of the living from the dead is greater than any division between nations or caste. While many transhumanists have proposed abolishing death going forward, Nikolai was nearly unique in proposing a retrospective abolition of death. Although a Christian himself, he thought, rather boldly, that it might be possible to resurrect everyone who had ever died using science. Without human intervention, salvation would be partial—only for good Christians, or perhaps only for

members of the Russian Orthodox church, but a mechanical salvation was possible. Such a salvation would not just restore all humans to life, but make that life eternal through the marvels of science.

If nothing else, what a sweet vision. There's the obvious, of course: for a hopeless romantic such as myself, Alexander and Hephæstion, Abelard and Heloise, Antinous and Hadrian, Andromache and Hector, Anaïs Nin and Henry Miller, whatever real couple the story of Apollo and Hyacinth was based on—and that's just couples with names starting with A & H. But far more important than these, nameless peasant 10,405,771,606 whose story you never heard, even though it was far more tragic. The approximately 5,000,000,000 dead of malaria. The roughly half of all children who never made it to adulthood. The lost and broken who lived a long life filled with ceaseless pain. Can you imagine how excited you'd feel if you thought for even a moment that you'd found some way to fix it all? All the jagged sheet of history with misery scrawled on it, folding into something beautiful. It's a holy thought—I would love to have met Nikolai. Indeed it's *just* possible that one day I shall.

What are the scientific prospects for this task? Before we get to that, we need to take a detour through philosophical theories of personal identity.

Personal identity

You step into a teleporter, it vaporizes you. A person qualitatively identical to yourself steps out of a machine somewhere else, with “your” “memories”, “your” personality, “your” body, etc. etc. There are two philosophical questions here:

- A) Have you survived? Is the creature that stepped out at the other end “you”?
- B) If you have not survived, is the outcome, from a self-interested perspective, i) as good as surviving, ii) better than ordinary death but worse than surviving or iii) as bad as dying in any other way?

If you think the answer is yes to A, or no to A but option i) to B, then you're in luck. The

Common Task might, from your point of view, resurrect the dead, (or as good as). If you answered otherwise, then the Common Task is unlikely to work, unless we can find some way to actually pluck the dead from the past. If you're interested in these topics, Google "philosophical theories of personal identity".

With that sorted, let's go on to "the science".

The possibility of a trace

Nikolai himself hoped that as we gained mastery over the physical world, we would be able to, based on some trace left by the dead, reconstruct them in body and mind. It's hard to say much on this, except that if it is possible, it would require—as best as I am aware—as yet unknown physical principles. Whether you think this makes it vanishingly unlikely, or reasonably plausible, is something of a matter of epistemic taste. As science has advanced, it is true that we have gained access to traces the ancients never would have imagined—DNA and carbon-14 dating, for example. We now could, in principle, reconstruct the bodies of some of the dead through cloning so long as we have their DNA. While this would not fulfill the great task, it is an example of the advance of science uncovering previously undreamt ways of reconstructing that which existed in the past. There is always the possibility of more such discoveries in the future. It is possible we will uncover some, as yet unknown, natural version of the Akashic records, although there seems no particular reason to hope so.

One trace is fleeing Earth at the speed of light—light. Light goes slower through some media than others and can be refracted, so in principle, it might be possible to capture the fleeing light without exceeding the cosmic speed limit c . This could then be used—again only in principle—to reconstruct events on the Earth's surface. In practice, there may not be enough information left, and even if there were enough theoretically, the engineering problem may be intractable even for galaxy-spanning super-intellec-

Nearly infinite simulation

Suppose that computing power turns out to be really plentiful. Maybe we can build computers from subatomic parts, for example. Now suppose we enter as constraints everything we know about the past and conduct simulations of the past, weaving endless quadrillion lives and creating numberless people. At the end of each of those lives we take the persons so created and put them in a digital afterlife. Eventually, for every person who has ever lived, one of those people is going to be arbitrarily similar to them. If you think that someone having had an arbitrarily similar life to yourself existing in the future counts as survival, you will have survived—congrats!

The process would likely be vastly more accurate for contemporary humans because the endless gigabytes of what is known about us means there are far fewer gaps to fill in with estimation. But while this may give you and your loved ones better odds, it's cold comfort for the long-cold nameless peasant 10,405,771,606, whose best approximation is liable to be far looser. Looming over all of this, of course, is the possibility that we are in a simulation ourselves. Whether that would make the task more or less likely, or whether it might already be underway, will remain open questions. I try not to think about this too much.

Time travel

One easy solution, were it possible, would be time travel. Most plausible conceptions of time travel developed in contemporary physics and philosophy of time suggest that it would be impossible to change the past. That would not necessarily foreclose on us going back and grabbing the data.

There is a reasonable review in the *Scientific American*: “According to current physical theory, is it possible for a human being to travel through time?” of the prospects of time travel by an expert targeted at a lay audience. The conclusion seems to be: “No.”

I asked my old PhD supervisor, a leading philosopher of time, and she gave the same answer. It's interesting to wonder if there might not be a way of grabbing data from the past without traveling back there per se.

Something we haven't yet imagined

Do you think we're near the end of discovery, or do you think that there are things as yet undreamt of in any philosophy that will one day be dreamt? Almost every human that has ever lived would be unable to understand options 1 to 3, so who is to say there isn't an option 4, 5 or 6? For more speculation, see *Resurrecting all humans ever lived as a technical problem* on Less Wrong, published since this article first came out for more discussion of possibilities. See also the subreddit Quantum Archaeology.

Summing up prospects:

To be honest, none of the above methods are especially persuasive to me, at least for all humans that have ever lived. I can conceive of something like option 2 working for humans who lived post the invention of social media, and for the relative bare handful of humans who left substantial documentary traces of themselves prior to this. Dan Simmons imagines something like this being used to recreate the poet Keats in the novel *Hyperion*. With those exceptions though, I'm not convinced the great task is feasible.

But it might be worth considering the pessimistic meta-induction. The pessimistic meta-induction is as follows: Since historically most of our best science has not turned out to be even approximately true, it seems probable that our contemporary best science will turn out to be not even approximately true. Personally I am skeptical of the pessimistic meta-induction in most areas of the philosophy of science, but a related argument which I call the optimistic technological meta-induction seems more plausible. Past attempts to define what would never be possible through technology were very often failures. This is especially true of technologies which we might see as

precursors for the great task. Cloning and the creation of life have both been achieved; projects for creating minds and superintelligences and achieving physical immortality are both underway and will surely be completed at some point if we don't wipe ourselves out. Most of those who have ever lived would not have thought these things possible for Mortals.

Technology so often surprises us, and that which we thought impossible happens so regularly, because, in the words of the IRA to Margaret Thatcher: "You have to be lucky every time, we only have to be lucky once". Out of all the harebrained schemes by humanity and its successor superintelligences, only one has to work for the great task—or any task—to be achieved. This is what makes betting against the possibility of anything—especially before we even know the fundamental laws of physics—so dangerous.

If you put a gun to my head and asked me to give you a credence I'd say there's only a small chance of this task being possible, but enough that I wouldn't bet with any great confidence against it.

Practical implications of the common task

There is nothing we can do to make the very unlikely resurrection of the dead more likely. Either it's possible at some unimaginable reach of technology or it isn't. About the only contribution we can make is fighting against humanity's extinction, and we should be doing that anyway.

There are lesser tasks which partially fulfill the great task but are time-sensitive and which we can make a contribution to. For example, ending involuntary death. If anyone takes this essay as a reason to aim at these less urgently, I will personally hunt you down and tell you off.

Maybe I'm setting myself up for heartache in the future, but I like to sometimes use the

common task as an organizing myth in my life. A sense of what would be the ideal outcome of everything, to measure and assess more feasible alternatives against. The role of an ethical-aesthetic organizing principle is difficult to explain, but it seems to help. Sometimes, when I'm at my bleakest, it becomes a reminder that no one has ever scientifically proven that everything won't turn out all right.

For a philosophy of the common task

Someone (maybe you!) could write a very good book considering the common task from a contemporary philosophical perspective. The common task raises questions in the philosophy of personal identity, time, physics, ethics and metaethics, religion and many more.

Indeed, you could teach a very good introductory philosophy course using the philosophy of the common task as a springboard. Time travel (theories of time), personal identity and persistence over time, the simulation argument (epistemology & scepticism), duties to the dead (utilitarianism against alternatives), the sceptical meta-induction (scientific realism and its rivals) etc.

Perhaps the most unique question is about the value of resurrection. Let us suppose that we can raise all ~100 billion people who have ever lived from the dead, but that we could instead use those resources to create, say, 200 billion new, joyous lives. Should we do the former or the latter? Do we have a duty to the dead to restore them to life if we can? Do we have a duty to the living not to leave them alienated from the dead? Do we have duties to the dead in a way that we don't have duties to the hypothetical persons we could bring into being?

And if we do start raising the dead, do we raise all of them? Presumably if we had the technology to do this, we could keep everyone safe from everyone else, but morally, does the world need Mussolini or Bundy back? Should they at least serve some kind of sentence before joining everyone else? These problems might seem absurd (they are,

really), but it's an interesting way to grapple with questions about the reason and purpose of punishment. The seeming absurdity of punishing anyone under these conditions is one of the reasons I believe that punishment can only be instrumentally good.

Aesthetics of the common task

There is an endless range of poems, t-shirts, mini-series, sculptures, novels, paintings, desktop backgrounds and radio plays waiting to be written on the common-task—or on the idea of a secular resurrection of the dead. At present I know of only two works, neither of which I have read. One of which, which I have forgotten the name of, tackles Nikolai's work explicitly, whereas the other is *Riverworld*.

Aesthetically the idea is almost megalomaniacal and difficult to pull off. It undercuts a central concept of much contemporary literature and art—death as an inescapable existential problem. Part of secular maturity is accepting the permanence of death, so the idea represents, in some sense, a return to a philosophical childhood. I feel it myself, even as I write this with unusual giddiness.

In the past I've suggested that post-scarcity worlds are very hard to write about because they remove many (though not all) of the obstacles that fuel narrative and that this leads authors to a kind of reflexive anti-utopianism, equating that which is bad for story with that which is bad simpliciter. A post-scarcity, post-resurrection world redoubles these problems. Add to this the human tendency to rationalise even involuntary death as a good thing to cope, and I can only suspect many authors and artists would instinctively oppose the great task. It will always find purchase among some, though. Some human problems do remain in a context without death or material scarcity, and while it would be very difficult to write a narrative about a context like that, I think it could be worthwhile.

Ultimately we have to grapple with religious aesthetics if we're going to try to represent

the common task. In a personal capacity, I find myself wondering if I am not trying to reconstruct the Christianity of my youth in a thinly secular context. While I don't believe, I can't help but dream of a glorious τετέλεσται, even if it makes me a sucker.

Fuck death.

Bibliographic note for “Oh Death, Where is the Antidote to Thy Sting”:

I had (inexcusably) forgotten an email I'd received from Damian Tatum that mentioned computer simulation as a strategy for resurrection. Although I had forgotten the email because it came during a busy period, I can only assume it influenced my thinking on the matter since the parallel between what he and I describe is strong. Alexi Turchin has also written an essay which covers very similar ground, though in a different way. You can find his essay here:

https://www.academia.edu/36998733/Classification_of_the_approaches_to_the_technological_resurrection

Perspectival fever: On being shot through with philosophical desire

I. Trapped behind my own eyes

I have a recurrent experience where something I've done, or something I am, appears to me to be better than I know it to really be. I pause and try to break through, to see it from the outside, but I can't.

When I write, what I've written often seems to me to be really good. Only I know it isn't, at least not usually, because the kind of "really good" I'm aiming for means moving people and changing minds. I have made over 170 posts, and done plenty of self-promotion. If what I was doing had the power to move a lot of people, I'd have a lot more than 100 daily readers by now. Tempting as it would be to write this off as market failure in the bazaar of ideas, I don't think it is.

I finally worked out the trick, though—the source of the illusion that makes it hard for me to see my writing as it really is.

The writing is the melody, and the harmony is my own mental state. Is it any wonder that those who only hear the melody alone do not enjoy it as much as I who can hear the joining of the melody and the harmony?

It reads so good to me because what I've written vibrates in harmony with what is in me at the time I wrote it. The insights feel fresh and powerful because they've just impressed themselves upon me. The metaphors seem choice because, almost by definition, I must have been in the mood for that metaphor when I cooked it. The only time I can see what my writing looks like to an outside observer is weeks after I've published it, when I reread it. In that moment I see it as sometimes pale, sometimes gaudy, sometimes obscure, sometimes basic, but never quite singing the harmonies I

recall. This is not because the notes have changed but because my mind no longer harmonizes with them.

Really, this is why all communication is destined to succeed, at best, partially. We communicate only the melody of ideas, but the harmony, the inner context that vibrated in time with that melody of thoughts, is not communicated.

I often feel that if I could just grasp what I'd written from a wholly different perspective, from many different perspectives, I'd be a long way to being a better writer. Don't misunderstand me, I have the ordinary capacities for empathy; what I crave is extraordinary capacities for empathy.

Could I reverse the process? Take a mediocre (though not bad) piece of text and think myself into the state the author was in when they wrote it, making my thoughts ring in harmony with it till it reads like a model of brilliance? I'm very interested to try. If you've written something you think is mediocre but which at the time felt brilliant, email me. I'll see if I can't bridge my way to where you were standing.

II. The fever

Writing is not the only time I have this experience.

When I was overweight and I looked in the mirror, my stomach seemed big, but my face never seemed pudgy. However, I knew it was pudgy from the testimony of others. But stranger still than that contradiction, I noticed that when I took pictures, my face seemed pudgier in the still image than in the mirror. I took thousands of selfies of myself over a decade, not out of narcissism, but because the discrepancy between the two was maddening. I wanted to understand by seeing myself as a stranger, to become an object for analysis.

It's not always a matter of my intuitive reaction being more positive than what I know to be warranted. When I suffer very bad fits of OCD—the kind of fits that make some psychologists suspect the disease is related to schizophrenia—I find myself maddened by a paradox. On some level I know it's all false, because it's all turned out to be false so many times before, yet I cannot fully see the world as one who thinks that it's false. It's not just a matter of feelings (aliefs); the bits don't click together right at the level of belief either. I try to see myself from the outside, but I can't escape my skin. As it once occurred to me in a daydream, there is a storm, there is a boat, I am the storm, I am the boat, and there is no way I can sail free.

I call this clash between what I know and feel, combined with a maddening desire to harmonise, perspectival fever. Reading what I've written, thinking through what I fear, and looking in the mirror—this is when it hits. I know something to be true, but I can't see it that way.

III. Do they know what they do?

A woman, who I will not name because she has already endured more than enough public shaming, wore blackface to a Halloween party. This was a confused attempt to parody Megyn Kelly's denial that there is anything racist about wearing blackface to a Halloween party. While her gesture was, by all accounts, intended to be anti-racist, it missed. As far as anyone can tell, she had no idea how poorly her choice of costume would be received until she turned up, at which point it was too late. Two years later (???), The Washington Post ran a story about her transgression, apparently in an effort to cover their ass about something.

I used to wonder if maybe most serious transgression and crime in our world was like this woman who somehow didn't realise she was going to get canned for blackface. People just sort of forget that a certain course of action is monstrous and illegal until it's too late. "I'm a murderer/rapist/mobster? Huh, well, I never thought of it like that, but now that you put it that way..."

Call the mental state of being unaware that you are about to do something transgressive when it should be obvious moral blindness. Anxiety about the possibility that we have suffered, or will, suffer moral blindness is pretty common—among the highly religious, in the anxious, shifting enclaves of this age and in various mental illnesses. So many people are afraid of fucking up and not realising till it's too late, but people don't often talk about it, in case they end up looking like a weirdo. In other words, people fear temporarily losing the ability to see their actions from the outside and breaking a rule. What is that fear if not another form of perspectival fever.

IV. Jiang

Perhaps it isn't surprising that someone with these strange derangements of mine would come to be obsessed with the following thought experiment.

Someone, call her Jiang, fell into a deep sleep, and woke up proclaiming that she had experienced all of it. All of human history. All hundred billion lives, or approximately three trillion years of experience. The weeping of Alexander as he saw there was nothing left to conquer. The passion and fear of the suffragette Emily Davison as she fell under that horse. The moment calculus first clicked for Newton. The plight of Pocahontas in England. Above all of these, though, the nameless and numberless of history and their unending days.

A day can be a long time; she has experienced over a quadrillion days. Jiang—and who knows how much of her remains after this experience, but we will assume she retains her identity—establishes her credentials. She then announces that she wishes to address the world. As she mounts the podium for a speech that will surely be watched by more people than any other in history, she opens her mouth and...

Do you think you know what Jiang will say? Not about everything, of course, but maybe you think you know what part of the message is? Stranger still, do you have a hunch that

Jiang will contradict some particular belief of yours? If so, how can you possibly justify your belief? Maybe I'm just uniquely thick, but I sometimes suspect I know what she might contradict me about.

Is the question of what Jiang would say even meaningful? The human mind as constituted isn't capable of processing that much data. Perhaps asking what Jiang would think, absent specifying how she would be modified to make it possible for her to process this total experience, is meaningless. Nonetheless I find myself longing, almost painfully, to know what the sum (product?) of all human experiences would be. Perspectival fever on a total scale.

V.Convocation

Another related question that entertains and torments me. Let's define a "grand convocation," as a hypothetical process in which all the people living in a polity were gathered to decide how a polity would be governed henceforth. Somehow there is allotted for them infinite time to speak and debate, and their capacity for boredom is removed. Each of them can address the whole as much as they like. Do you find yourself fascinated by the unknowable question of what they would decide? What if we vary it—for example, by binding them all to truth in their deliberations, or by greatly enhancing their intelligence. (Surely you must relish the thought of what they might sweep away?)

Or what about a grand convocation of Jians? What if every person in the assembly experienced the life of each other person? All approximately 330 million Americans (or substitute any other state, or substitute the whole of humanity) living the lives of the approximately 330 million Americans, on top of their own, and then—and only then—hammering out a consensus on how the US should proceed.

VI. To air myself out

Here's another philosophical fantasy for you, this one more exhibitionist than voyeuristic. Have you ever dreamed of presenting your whole self to others? Of giving them a copy of a biography of yourself written by God or a Culture Mind or something? Something that tells the whole story? Or maybe just having the magical ability to make others know you are telling the truth when you are, in fact, telling the truth—so you could tell them everything and answer their questions with complete confidence they would believe you?

Is it a fantasy of forgiveness? Of letting others see you have done wrong, but showing them so much of yourself that they can't help but see it in context? Or is it a fantasy of connection? Of releasing the ache of a loneliness so deep you had forgotten it was there?

It's all of these perhaps, but it's also perspectival fever—the desire to be seen through other eyes is, in part, the desire to see yourself through other eyes.

On Klutzes

On Chapo the other day, Amber A'Lee Frost raised an interesting point. Regarding the novels of John Steinbeck, she suggested that he was feeding into a national conversation happening at the time he was writing. A conversation about what to do about people who just aren't very good at things, but in a way not generally recognised by the welfare state. We will call these people klutzes. Klutzes are not simply unlucky in the labour market; rather, their talents make them poorly suited for it. Klutzes may or may not have disabilities, but their disabilities are not of the sort that can be used to make a case for social assistance under the current rules.

A personal essay tack is required here. Klutzes are a topic of great interest to me because

I consider myself a klutz. I'm dyspraxic physically—I have poor fine motor coordination. Cognitively, I suffer from an executive function disorder that makes me chronically absentminded, and while my spatial reasoning can work, it works very slowly. These factors led to me being fired from my first ever job as a kitchen hand by my own father (he's a wonderful dad, I was just a really, really bad dishwasher). Compounding these weaknesses, I have periodic bouts of severe OCD that can leave me suicidally depressed for up to a month at random. I have a certain degree of charisma and customer service skills, but nothing so truly exceptional as to overcome these handicaps. My skills could be summed as:

1. A decent but unexceptional work ethic
2. Reasonably well mannered and understanding, though with a somewhat off-putting tendency towards eccentricity
3. A capacity for research (though with a tendency to distractibility)
4. Writing skills
5. The ability to walk very long distances.

There are certain jobs which fit this profile of skills, but unfortunately they're extremely competitive. Consider, for example, being a writer. I have made pitches to many institutions and publications, including the People's Policy Project, Current Affairs, The Guardian, New Matilda and The Conversation, and not gotten a reply. At the risk of sounding bitter (and I am) I've been told that a lot of media organisations aren't interested in cold pitches. If you haven't already got a portfolio of published work, you need to know someone—or so the story goes. The dire odds of me making a living in media are further lengthened because: I'm eleven shades too left for the centrist establishment which holds 99% of the money in media and—this essay excepted—I'm generally averse to personalising my writing or marketing aspects of my life and identity. No doubt I could overcome these barriers were I a genius. I'm not a genius, and I shouldn't have to be. I'm a failson, but not the son of anyone important.

I have considered a lot of options. Onlyfans? I'm not ugly, but I'm a little too chubby—and even if I lost the weight I'm nothing special. Patreon? I don't have a large enough audience yet (I doubt I ever will). Starting a business? Most of them fail, and with my absentmindedness I'm more likely than most to fuck it up. Maybe you're thinking that if you had my skill set you could do very well for yourself—that I simply lack a certain can-do attitude. Perhaps you're right. If so, consider “positive mindset” as just another thing I'm a klutz at.

The result is that, while completing a PhD, I work a job that I am unsuited to. I hang onto it by the fingernails—through people skills and a contract that makes me very hard to fire. Once I finish the PhD and my scholarship money dries up, I'll make a -probably doomed—attempt to get an academic job. When that fails I don't know what I'll do next. The klutz is mostly invisible in our culture. The few exceptions are klutzes who manage to find some special niche for themselves. The absentminded professor, for example—a trope that some people have told me I fit. Really, such quasi-klutzes are the lucky ones. The real down in the dirt klutzes, the ones that have it very hard, don't have any marketable exceptional strengths to offset their weaknesses. They turn up in our culture

from time to time. For example, I don't know a lot about incels, but if my friend and host of the podcast *The Conditional Release Program*, Joel Hill, is right, many incels are klutzes with the additional disadvantage of being conventionally unattractive and hating women.

None of this is to say that klutzes lack talents altogether; they just lack the marketable ones.

As I mentioned, I am one of the lucky ones because my klutziness is incomplete. I'm just organised enough to fake it and hold down a job. While my parents are not rich, they are very generous and supportive. I suspect I would have spent at least a little time homeless and couchsurfing with friends if my parents weren't supportive. If I were charmless or just a little bit more disorganized still, my parents would have to stretch their generosity further.

I believe there's a lot of us klutzes around, though I can't prove it. There is certainly an outsized number of partial klutzes like me in academia.

Probably, since a young age, many klutzes have been told that they are lazy, because a lot of the time that's what being a klutz looks like from the outside—e.g. if you have an average or even somewhat above-average work ethic and are a klutz, you will look lazy to others, because you get through work more slowly. “He couldn't possibly have forgotten again, he must just be too lazy to do it.” “He couldn't possibly have broken something again just by clumsiness—he must not be paying attention”.

What I'm trying to do here is articulate a new group—to impose a new condition of being into your consciousness. The klutz. I'm not going to propose specific solutions here because that isn't the point—it is not beyond the wit of sapiens to provide a decent life

and meaningful opportunities for contribution to klutzes. It's a matter of will. Instead I am doing what I most hate: awareness raising.

So, conditional on us not forgetting the time or location, let us rally. Upon our badly homemade banners, let us raise the sign *omni homini habeat valorem*, which, if my amateurish Latin translation is correct, means "every human has value" (and if it is not correct, is that not much the better?). Let us fight for a dignity not conditional on technique or power. Arise, my broke and broken siblings! Let us chant, "useless, not worthless".

Existential tragedies—a partial list of the fundamental complaints of being a person.

We all know there is a suite of “existential” tragedies inherent to human existence. Defining the concept precisely is impossible, but I would say roughly that an existential tragedy is a tragedy that arises from very basic and universal, or near-universal, aspects of our experience. I thought it could be interesting to list them. Let me know if you can think of any others:

1. The inevitability of one’s own death
2. The inevitability of the death of those one knows and loves
3. The probable unknowability of many important cosmic truths
4. The apparent existence of meaningless suffering serving no higher purpose
5. The inherent trade-offs around what can fit into a single life
6. The possibility of unrequited love (and I certainly don’t just mean romantic love—unrequited familial love is usually worse)
7. Our total lack of control over the most important factor of our lives—the circumstances of our birth
8. The unequal and random distribution of talent
9. Our inability to consistently embody even our own idea of the good
10. The erosion of treasured (or simply important) memories by time
11. The erosion of vitality and beauty by time
12. The unknowability of the full results of our actions
13. The privacy of experience even when we wish it were otherwise
14. The frustration of words that can never fully convey what we mean
15. The perpetual possibility of being disbelieved, even when we are speaking important truths about ourselves and our lives
16. The possibility of being wracked by want for something which is impossible—not merely improbable but totally impossible.

17. The irrevocability of our actions—in a few lucky cases we might be able to prevent or fix all the harm we caused, but even then, this does not undo the action itself

Try to always be kind because you never know when you're incompetent

We are rarely one step from disaster. Most really awful outcomes require at least two things to go wrong. Often, but not always, those two things are being malicious and stupid at the same time.

Hanlon's razor says that you should never attribute to malice that which you can attribute to incompetence. It's not a bad way to think, but it's misleading if you take it too far, because most disasters are caused by a bit of both. For example, a wildfire department is under-resourced because the state just doesn't care enough, and the resources are misallocated within the department because the state is incompetent. A police officer is too incompetent to tell that the suspect is choking to death, and doesn't care about the person enough to stop simply because they are inflicting severe pain.

The good news is that you can short-circuit the synergy between malice and incompetence and often prevent the most disastrous consequences of your actions by trying to be a nice person. It's much easier to stop yourself from being mean than it is to stop yourself from being stupid—you usually know when you're being mean, but you usually don't know when you're being stupid. Most of the time you also don't have to be exceptionally kind to avoid disaster either, just ordinarily decent.

I worked in the Emergency department once in the graveyard shift, as an administration officer. One of my tasks was to buzz the nurse to let them know that a patient had arrived. Since the staff was so small, they couldn't spare a nurse permanently at the desk during the night.

We had a lot of patients who came in for absurd reasons. Sometimes it caused big problems and slowed down care for those who really needed it. Mysterious aches and pains were common, and often the patient would see the triage nurse, wait several

hours, see the doctor, and just get referred on elsewhere, because their problem was non-urgent. I know for a fact that other admin staff—not all, but some—used to make their irritation with these patients plain on arrival.

One night we had a guy come in one day saying his balls hurt. He wasn't in severe agony or anything, they were just aching. Mentally I rolled my eyes. However, I believe in being a nice guy, so externally I certainly did not roll my eyes. Instead I gave him a comforting smile and immediately called the nurse. I expected she would have a quick look and suggest he come back in the morning, as had happened to a couple of other patients that night.

Instead she immediately assigned him just about the most urgent triage category you can get with your heart still working and your limbs still attached. Afterwards I asked about him, and she explained he had a suspected testicular torsion and that the only safe way to treat it was immediately.

Had I rolled my eyes or questioned his coming in, as other admin staff sometimes did, I might have left the poor man a eunuch. It didn't take a heap of kindness to avoid it (I have no delusions about being a kind person), just a tincture of patience.

The moral of the story is that because being wrong feels exactly like being right, you're almost always better off being nice. Kindness covers a multitude of incompetencies, including incompetence you didn't even realise you had.

Brief Reflections

Having read Manguso's *300 Arguments*, I was tempted to attempt a few aphorisms of my own:

- The difficulty with writing aphorisms is that you've got to find a rare enough insight not to be trivial, but self-evident enough that merely stating it is persuasive.
- Aphorisms are innately the most egotistical form for this reason. "I don't even have to give you substantial arguments to be worth reading". I apologise for that. This feature of aphoristic writing explains Twitter.
- If Twitter is a game, the parameters of that game mean that the vast majority of its players must be losing.
- Creativity is not about turning off the filter. It's about turning a simultaneous process into a sequential process. Normally we filter our ideas at the same time we make them. By contrast, in a creative mode of thought, we make a reiterating two-step process of creating, then filtering, creating, then filtering... It's like the tug of war between the gods and demons over the naga serpent which churned the ocean of milk.
- It's very hard to find a method for individually working out whether a new thought is A) an original (or at least rare) insight or B) something everyone else has already figured out except you. This is why people in the business of thinking up original thoughts are all paranoid about being laughed at, egomaniacs, or both.
- The furtiveness of a bad conscience—its exact balance of shame and guilt and other fascinating parameters—is the secret master of the world. Secret because it's self-occluding. Definitionally people don't want to talk about it. This means its real influence is assuredly manyfold times larger than its apparent influence.
- Guilt's a pretty dark emotion. It almost invariably shades into grief that I have to deal with the emotional consequences of having done wrong rather than

grief that I have done wrong. The line between guilt and shame is also worryingly thin, as is the line between guilt and fear of punishment. Often the only thing more reprehensible than feeling guilt is not feeling it.

- One of the disturbing aspects of how the media has chosen to cover #metoo is that it is covered as a sex scandal. This means a degree of the interest is prurient. I remember a case broke about a footballer accused of rape in my country. The tabloid in question chose for its accompanying image the footballer in question in a pair of speedos.
- The best argument for loving everyone is that you can't fully love anyone unless you love everyone. We all contain each other's flaws and capacities, even if only in fragments.
- It's simple enough to get statistics on what percentages of people affiliate with religion. What I want is statistics on what percentage of people believe in god in the same matter-of-fact way they believe in Belgium. I want to know how this has changed over time. I suspect it's more dramatic.
- It's interesting that it's Astrology—rather than Tarot— which is undergoing a revival, despite the frankly superior aesthetics of Tarot. One suspects that the role of astrology in assigning identities is part of the attraction, much like people love the Myers Briggs for giving out labels. But equally important is its allowing you to label others, in a fragmenting world.

The questions that haunt me at 3 in the morning

I think that everyone has unanswered questions that bug them. Recently though, I've been making a point of jotting them down. I've begun to realize how many of these thoughts are reoccurring, and that these constellations of questions define my mental life almost as much as my beliefs.

You may think that you know the answer to some of these questions—you may well be right. You may think that the answers to some of these questions are obvious. Here I would have to disagree with you. There are times in my life when I would have agreed with that about some of them, but I've become less and less confident of them over time.

Psychology

- To what degree are people motivated by sadism in everyday and political life?
- Is self report a reasonable measure for variables like personality and happiness?
- What is the optimal amount of caffeine to ingest? How does this vary by life you are trying to live.
- Sovereign citizens believe both A) That the legal system is run by incredibly evil people entrenched in power B) That if they can just say the right sequences of words, those people will be compelled by the rightness of their arguments to let them off various crimes, civil liabilities etc. Obviously this combination of beliefs is irrational—that's not in dispute—but what about it makes it so compelling? Shouldn't the factors that make you feel the legal system is incredibly evil also make you feel like you're not going to be able to sway them just by making a strong argument? Why is this seemingly contradictory combination of beliefs such a potent attractor?
- Would the Milgram & Stanford Prison experiments replicate if run properly today?

- What the hell is going on with the replication crisis? A lot of the failures to replicate are in really simple experimental designs. How much of it is outright fraud? How much of it is pure chance and the file drawer effect? How much is ad hoc analysis and statistical fishing? How much is the participants giving the experimenters the “expected result”?
- How can we make trying to replicate experiments an honourable and attractive path given the structural incentives of academia?

Criminology

- What percentage of convicted incarcerated people are innocent?
- What would be an acceptable false positive rate in a just criminal law system? How does this apply to crimes which can be extremely destructive, but which are often by their nature extremely difficult to prove “beyond reasonable doubt”.
- Under what conditions -if any at all-, and in what ways -if any at all-, should society informally punish people for whom a criminal conviction is likely impossible?
- Are people who have done horrific things (rape, premeditated murder) generally all round bad people? Or is human character contradictory in such a way as to allow at least a significant minority of such people to be, despite it all, good or at least average or not far below in most other areas of life? This may sound absurd to some readers, but a lot of anecdotal evidence tends to suggest it might be true. What is going on here? To what degree do people have consistent moral character?

Misc human sciences

- To what degree are there political or social ideas which, even now, if someone dreamed them up and went to some modest effort to promulgate them would transform the world? Are such remaining unthought ideas relatively few, and

their effects mostly modest? Relatively many, and quite a few with great effect? Some other combination? The real underlying question here—to what extent is theorizing about political and social issues an effective strategy for changing the world?

- What would be the social effects of a working lie detector?
- Similar to the question about a lie detector. A wizard waves a wand. From that point onward, no one can lie. Does society A) collapse B) get much worse on net but continue C) continue with surprisingly little change D) get much better on net E) become a utopia?
- Is it possible to throw down the tyrant irony and return sincerity to her rightful throne?

Politics

- Our period is defined by greater political divergence on the basis of age than ever before seen since we started taking polls on this sort of thing. Preliminary evidence suggests so far that this is a cohort effect not an age effect auguring that the younger generations will not age out of their opinions as they get older—at least not automatically. Will this trend continue?
- To what extent are people who make comically bad takes about politics acting in bad faith—versus stupid or self deceiving? Especially those with the expertise in politics to know better? Let me give examples from both sides of politics. I recently saw a senator complaining about raising the minimum wage to 15 dollars partly on the grounds that when he was young he was paid only 6 dollars. Adjusting for inflation, this turns out to be 24 dollars in today's money. On the other side of politics, I recently saw a news station try to insinuate that privilege is the reason we are paying for the lawyers of many of the Capitol Hill rioters. In cases like this—awesomely stupid takes by people with adequate education to know better—are people A) Lying B) Bullshitting in Frankfurter's sense or C) engaged in feats of incredible self deception or D) apocalyptically stupid This might sound more like a

complaint than a question, but I mean it sincerely as a question—what is typically going on here?

- Why aren't democratic politics more effective? There are certain issues on which the average member of the public disagrees greatly with the average politician, and yet there is very little movement. Obviously money in politics has something to do with it, but how does money exert its influence exactly? It's easy to get blasé about this replacing incredulity with a faux wise cynicism, but if you stop and think, it's weird that 70% of the public can strongly support something, yet it be considered a fringe position among politicians. How much of it is caused by monopoly power wielded by political parties due to existing voting systems? How much of it is due to rational or irrational voter ignorance? How much of it is caused by voters deceiving themselves about what they believe—with their true sentiments closer to those of politicians? How much of it is due to the action of the media? What forms of action by the media (and other ideological apparatus) count here?
- What on earth happened sometime in the mid seventies? The wage stagnation, the incarceration spike, the union membership decline...
- Is neoliberalism a useful concept for understanding this phase of economic and political life that began sometime in the mid 70's? Was it ever useful and is it still useful?

Philosophy

- In philosophy and several other disciplines it seems like people were having more “big ideas” before about 1970. The period from Frege to the tail end of logical positivism and ordinary language philosophy – or maybe even till 1980—seems to have been very fecund. Is this an illusion of hindsight? Or is it possible that we've exhausted the “low hanging fruit” in certain disciplines, and the remaining brilliant theories and questions are harder to find?
- What the hell are qualia? How do they fit into the world?
- There are many positions which I think are good “in an aspirational

sense”—as organizing myths and sources of inspiration—but would not work in practice—at least with the current level of technology. These include communism, prison abolition, and to a certain extent, in some moods, anarchism. Is this meaningful in any way, or is it just a glorified way of saying “I wish things were different”—a vacuous waste of energy?

- Is there a compact definition of what it means for a person’s life to go well for them which doesn’t imply that we should do something absurd—like tile the universe with people enjoying their best moment over and over again?
- A lot of questions in philosophy seem like semantic wrangling. Often undergraduates will claim that this or that topic is just semantic wrangling. Why is it so rare for professional philosophers to openly embrace the view that a lot of philosophy is semantic wrangling? One could then argue that’s nothing to be ashamed of because some sorts of semantic wrangling and surveys of the semantic landscape are important.
- Is consequentialism self-effacing? Would we consequentialists be well advised not to hope that consequentialism becomes widely appreciated common-sense because this might have various bad consequences (undermining integrity, inspiring a certain callousness, allowing people to rationalize their preferred course of action)?

Economics

- Most actual businesses don’t face rising marginal costs—instead marginal costs are typically either falling or constant. How should neoclassical economics be transformed in light of this?
- What lies at the bottom of the Cambridge Capital Controversy well?
- Why do we keep using economic models even when they’re rubbish and there’s very little evidence they’re informative rubbish?
- Why don’t we talk more about the above three questions?

Art

- An argument can be made that poetry was once the most important of the arts, yet now it is effectively dead!?! Is it because poets became too focused on avant gardeism and not enough on classical technique (esp rhyme & meter)? Is it because singing -when done well—is strictly superior to poetry for most people, and technological advances make high quality and/or catchy singing always available? I tend to favour the second explanation, because the first explanation implies that poets are just “leaving free money on the ground” by not returning to classical technique. Nevertheless, this is a great artistic and cultural mystery. Even the fact that there is not more discussion of the disappearance of poetry is a mystery in itself.
- Why is so little fantasy and science fiction writing literary in the capital L sense? Why has no one written *Love in the Time of Cholera* except with fae? *Ulysses* with familiar spirits? Obviously there’s magical realism, but I’m surprised more people haven’t tried a blend of literary writing and genre fantasy. Maybe fantasy writers are too smart to fall into that trap?

Sufjan Stevens

- In the song “The Mystery of Love” the narrator says that they are “like Hephaestion who died, Alexander’s lover”, but it seems they are actually the opposite of Hephaestion, who, after all, died still the beloved of Alexander. In what sense are they “like Hephaestion”.
- What does Gideon, the biblical judge, have to do with a breakup?
- Why Ursa Major?
- What sense to make of “The Ascension” the title track of the eponymous album? Who is Cordelia? What is a chariot hallucination?
- How to understand the dialectic of a singer whose songs are very personal and whose life is very private.

AI safety related

- How much smarter than a human would you have to be to trigger a singularity?
- Imagine a person started off in good health in a first world country with an apartment, internet access and \$5000. How much smarter than a human would they have to be to take over the world in one year?

Medical

- It seems like we should be able to greatly reduce the length of time it takes to become a medical specialist. Indeed the length of time required smacks of industry protectionism & regulatory capture—especially when one considers that the various colleges self-regulate who can and can't access the relevant titles. Yet these lengths of time seem pretty consistent around the world. Is there anyway we could safely slash these times by a third or even a half (e.g. six years to become a psychiatrist out of high-school)? Is the specialty structure itself, where everyone learns to become a doctor and then specializes, the best way to handle medicine?
- Linked to the above—medicine generally has been more resistant to proletarianisation than any of the other professions (law in particular has been smashed, pharmacy even more so). Could medicine be proletarianised and doctors made just another type of worker? Should it be?

Personal

- How can I live an authentic life, true to what I actually think, without being wildly ineffective and getting people offside? Presumably the answer is a compromise where I try to keep the lion's share of both my effectiveness and authenticity by sacrificing a bit of both, but isn't sacrificing a bit of one's authenticity for strategic reasons massively inauthentic?

Chess

- How close are we to the ceiling of performance in chess?
- Consider a being that can see all possible game trees in chess, and in particular knows for each move whether it keeps the current game state, (win for white with best play, win for black with best play, draw with best play) or pushes it into a worse game state for its side. It is playing against a superb human grandmaster. Presumably the starting position of the game is a draw with best play. The being then picks from those moves which don't "slip" the game state into something worse for it. Is there a relatively simple way to describe an excellent strategy for this chess super being to increase the likelihood of its human opponent slipping up and making a move which throws the game into a win state for the chess super being? of such a chess super being, E.g. one option, although I think it doesn't work, is "choose those actions which increase the total number of losing options for my opponent, even if she picks the shortest path to a draw".

Autopsy on a dream

0. You're standing at the bus stop as the bus comes, looking for your bus pass. You search pockets and wallet compartments with frustration. At a particular moment, this frustration flips into something else. Previously you wondered how long it would take you to find it, and how much of a fool you must look flailing around for it, but as of that instant, the nature of your problems transforms. You now know you won't find the pass. Of course, it doesn't stop you from searching for the pass, but your search has a fully different character. We need a word for the moment of this flip, the figure-ground reversal of hope and defeat.

1. Many years ago: I am in professor Paul Griffith's office. I have just told him that I try to write two thousand words a day. "That's too ambitious" he replies. "But Bertrand Russell wrote two thousand words a day", I reply. He looks at me with gentle humor. There is no malice in his words, but there is certainty "You are not Bertrand Russell".

2. In Terry Pratchett's discworld novels, it is remarked that million to one-shots almost always work. In astonishing defiance of reason, kindness and prudence we teach our hearts the same thing. I'm sitting in my office, newly 33 when, perhaps for the first time, I believe that I'm not going to make it as a writer. I have long believed that I am not going to make it, I'm not a fool, but mostly I did not believe it. I knew the odds were too long, but I felt I would succeed anyway. We spoil our own hearts, and then we beat them.

3. Ages. Immanuel Kant, *Critique of Pure Reason*, 57. David Hume, *Treatise on Human Nature*, 28. Sufjan Stevens, *Illnoise*, 30. Leonard Cohen *Hallelujah*, 50. Jesus Christ, *The Sermon on the Mount*, 32. Karl Marx, *The Communist Manifesto* (yes that is his best work), 30. Have you spent hours on Wikipedia doing this too? Googling late bloomers and the like? It's one of those weirdly specific things that turns out to be a very common experience—like worrying that your dead relatives are watching you masturbate.

4. I look down at a referral for a patient in front of me. She is 57 years old. Her conditions include “urinary & bowel incontinence, schizophrenia, PTSD, agoraphobia, anorexia nervosa” and a dozen more. Her medication list has no less than 22 items and is evenly divided between Tolkien elf names, and what I can only assume are the names of dead-dread gods (why would you call it Cephalexin???) She has come to us for treatment of her osteoarthritis of the spine, accentuated by her weight (approximately 400 pounds). I call her, determined to get her into our clinic at once. I am even convinced that maybe getting her booked in is the reason I was born. I get her answering machine twice. I finally speak to her, she explains we can't schedule the appointment too close to the anniversary of her mother's death, since grief will distract her. My heart is breaking, but it's nothing against the shivers in her voice. The only happy emotion I can find is pride in her for staying alive. I book her in and finish up the call. Then I get back to worrying about the fact that I'll never be an accomplished author, vanitas, vanitas.

5. A few months before this, I am browsing the internet when I see some people talking about my blog in a forum. One of them mentions that it had “saved his life more than once” referring, I think, to my posts about OCD. I am confused and humbled. I am with friends, and I show them what's on my laptop screen like I am a kid with a puppy and show and tell. I have forgotten, more or less, by evening.

6. I hate myself for the kind of myopia typified in 4 & 5, but I'm far from the first to observe that hating yourself is just another form of vanity. Musing over the fact that hating yourself is just another form of vanity is also another form of vanity, and reflecting on that is another level of vanity again, so on and so forth up a conceited ladder to a pointless heaven of self-absorption. Nothing for it but to stop worrying about how vain you are and try and do some good.

7. Even at 33 I am already noticeably dumber in some respects than I was five years ago.

Much worse, I would estimate time goes 40% faster, and that's at minimum.

8. As I write this, I think about how a gifted prose stylist would regard these scratchings, I can't imagine they'd see them well. "Well fuck them", I resolve, "They're the freak, not Me".

9. The thirst to create drives out the thirst to know, and the thirst for fame drives out the thirst to create. Once, before all these when I was so young as to scarcely remember, there must have been the thirst to see.

10. "Unhappy is the land that needs heroes". Well sure, but imagine, as is sometimes posited in fiction, that there was a genuinely final confrontation with evil. Good, represented by some hero or small band of heroes, triumphed. After that life was to be wine and cake, but with no glory. Can you imagine how bitterly so many people living in that world would resent our hero? All the more bitterly, because bitterness is most bitter when it can't show itself. All the more bitterly still again because bitterness is most bitter when all morality and convention says it should instead be gratitude.

11. I think it's interesting to imagine what the life of our hero -the last hero—in the above situation would be like. For example, he might spend the rest of his days walking as if the ground might fall from underneath him, in fearful recognition that many people must be seeking an opportunity to reverse the reputation they had grudgingly given him. On the other hand, perhaps she would enjoy the pleasure of being a tyrant as well as a hero. She might gloat, internally, that she was the one who had prohibited heroism, while greedily drinking the remaining supply. I hope not, but then again, heroes often aren't good people.

Confessions: A psychological and intellectual autobiography

I confess to almighty God that I have sinned through my own fault in what I have done and what I have failed to do.

-The *Confiteor*, as I have remembered it

Everyone seems to want to tell their story- from opinion columnists to recipe bloggers. We live in a period of writing as-personal-narrative. I thought it would be interesting, fun, and above all expeditious to tell you the whole thing at once. Why laze about, making hints and insinuations, when I could just spill? I'm aware this is a narcissistic exercise, and that kept me from doing anything like this for a long time, but finally, I figured "Hey, it's not going to hurt anyone, I haven't seen this done before and it could be interesting".

You will learn little about the external events of my life here, but you will learn more about who I am than any but my closest friends know. I promise you also that everything in here is true. I've reread it to make sure not only that it is true, but even that nothing in it is misleading or will misdirect you. I have cut back on some explicit details, both because I'm afraid of awkwardness for us both, and because I am afraid simpliciter. However, I have not made any omissions that would even brush against dishonesty. Long-time readers will notice that a lot of stuff I mention here has been covered in its own essay at some point.

Core psychopathology

- My greatest fear, as I have confessed elsewhere, is harming other people. This is how my Obsessive Compulsive Disorder manifests. When I was younger I experienced this in a noble way- fear of harming others. As I got older it became less noble. It became a fear of hurting someone and then suffering reputational or legal consequences.
- As a result of this fear of harming others unintentionally, I have a deep fear of *not being known*. Of people thinking I am evil and cruel because I have harmed

someone accidentally. The worst possible thing that could happen to me would be me getting framed as a serial killer or something like that. I don't need to be universally loved, but I have a deep need not to be reviled.

- Because my greatest fear is unintentionally harming others and facing the consequences, I have sympathy for people who are abhorred. Rightly or wrongly I think “there but for the grace of God go I”. Or as Sufjan Stevens put it *and in my best behavior, I am really just like him* (John Wayne Gacy Jr).
- A lot of my OCD focuses on what I call a Sword of Damocles complex. I think that something or other I did years ago, maybe something I've forgotten or didn't even realize was a problem at the time, will come to light and destroy me. No one will believe my protestations that I didn't mean it, or that it was an accident.
- I have a theory that a person's greatest fantasy is often a reflection of their greatest fears, so we come now to my greatest fantasy. My greatest fantasy would be to receive something like the following trade. I will be rendered unable to lie, but everyone will know that I am unable to lie. People would know who I am because I would tell them, and they would know that I am telling the truth. In the spirit of complete honesty- by greatest fantasy here I mean something like “the most integral to who I am”, not the most attention capturing- that one would probably involve Henry Cavill doing some rather nasty things to me that I'm sure he's too much of a (straight) gentleman to do [does he even have a basement??].
- I'm fascinated by lie detectors as a possible means to fulfilling my fantasy. Even though I know it would be dangerous, I would very excited to learn that someone had made a working lie detector. I know that having a working lie detector would probably be nothing like I think it would be. Nonetheless, I still find myself Googling technologies that I see as related to lie detection, for example, brain-computer interfaces, machine learning-based thought reading approaches, and others.

- I am amazed at how little interest there is in this topic from transhumanists etc because, in my more fantastical moments, I think that the creation of a lie detector might be an end condition for history. That is to say, lie detection may be a technology that would bring about a state change in society, a horizon past which we cannot see. I think about the changes- to politics, to business, to law, to international diplomacy and so on that would result. I suspect the cumulative effect would be either utopia or a dystopia.
- Related to the above about lie detectors, I suspect that 5-10% of the population are ruining it for the rest of us. So many social designs and arrangements become viable if you can rule out the possibility that a selfish bastard will interact with them. So many dialogues become possible if you know the person you are talking to isn't secretly a dickhead. Dickheads are usually easy to spot, but nonetheless, the possibility that someone *might* be a secret dickhead leads to a kind of everyday Hobbesian trap. The feeling that a handful of dickheads ruin things, and the inclination towards mercy even towards the worst people, contort together within me.

Miscellaneous psychology

- A lot of people say things like "X is my spirit animal". I would not think about it in those terms, but I have a very strong identification with bears, particularly grizzly bears. However strong you think this mental identification is, you're probably undershooting. I have had this sense of connection since I was a young child.
- I have watched maybe ten to fifteen movies in my adult life. I recall watching only one TV series (True Blood)
- I hate my body. I'm a little overweight, and that certainly doesn't help, but it goes much deeper than that. Even if I were at 8% body fat I'd still be bothered by my lumbering frame. I'm 6'3" and broad-shouldered even for that. My face is handsome rather than pretty. I'm a twink trapped in a (semi) hunk's body, but

let's be real, I mean I am a sub trapped in a dom's body. Faced with few options, I decided to give people what I wanted so much myself. I won't get into the details about how this interacts with harm OCD and sexuality, suffice to say the complexities are complex. I'm sorry if this much is too much, but I felt in an intimate psychological portrait I probably had to say something about sexuality. I'm sorry also if my problems in this area seem trivial compared to yours, in some cosmic sense I'm sure they are, but they matter to me.

- I am talking a lot here about my foibles and weaknesses, but I don't want to give you the wrong idea. On the surface, I'm pretty healthy, a bit melancholy but mostly stable. Only my close friends know that I'm filled with turmoil. I've never self-harmed or attempted suicide. Maybe once every 3 years on average I'll have a breakdown and my family and I will consider getting me hospitalized, reject the idea, and then I'll get better over a month or two.
- I'm a great believer in tossing around ideas without being sure they're true. This is because, in my view, we should view our aim as trying to help society reach the truth in expectation, rather than reaching truth ourselves. I think things that are often seen as epistemically detrimental- like an individual latching onto an odd idea and seeking to prove it, rather than test it- can actually be healthy from the epistemic perspective of the whole.
- I don't think I've ever been envious, in the sense of disliking someone for their beauty, or intellect or anything like that. I wonder if I am unusual in this regard, or whether envy is much rarer than certain moralists have claimed. When you squint at it, envy does sort of seem like a self-serving myth "Oh he hates me because I'm smart".
- My favorite song is *The Mystery of Love*. My favorite poem is *The Four Quartets*. My favorite book is *The Communist Manifesto*.
- The error bars on my own sense of moral standing, relative to the rest of the species, are huge. Yet I feel that the fact I even care puts a floor on how bad I might be (for a wicked enough person wouldn't mind) and a ceiling on how

good I might be (for a truly good person wouldn't care how good they are, thinking only of trying to do better).

- There's a certain public figure I have a crush on (TBC- not Henry Cavill-that one's mostly a joke). For very sensible reasons, despite it being completely possible, I have never made a move on this guy by, for example, writing to him, so it's clear I have no real interest in attempting to seduce him. Yet to amuse myself, I tell this mostly-joking, a mental story about how various things I'm doing are to try and get his attention. Building a public profile. Working on stuff we would both find interesting etc. It adds this kind of fairytale romantic comedy element to my life. Again, to be clear, none of this is real. If I wanted to shoot my shot, I would have written him a letter about shared artistic projects, I haven't done that for ethical reasons, but playing this weird mental game in my own head is somehow sustaining to me.

Politics

- Politics is a perpetual struggle between the strong and the weak. Spelled out in those terms, how could anyone help but choose the weak? After all, people have declining marginal welfare in things/status, and tautologically, the weak have fewer things/less status, so the weak getting more of what they want will usually be more welfare increasing than the strong getting more of what they want.
- Let's define a political loonie. A political loonie is someone who is driven to be insufferable by their politics, prone to constant outbursts of emotion, etc., etc. Loonie is not synonymous with distance from the center- I'm on the far left, but I'd like to think I'm not a loonie. Centrist loonies certainly exist.
- If you go on Twitter, and you look at the loonies, you'll see two types, both defined by their inability to tolerate the contemporary world. The right loonies can't stand how ugly it is. They have a very narrow range of aesthetic acceptability, and they have trouble outside that. They are disgusted easily- for

example by the thought of sleeping with a woman who has slept with someone else, by unconventional aesthetics and lifestyles, and so on. Look at their propaganda, half of it is sharing ads (so idealizations to begin with) from the past and saying “look at what we lost”, as if the world had ever been so “fit for consumption”. They hate the fact they don’t live in a beautiful storybook.

- The left loonies on Twitter are also can’t handle the modern world, but they are filled with rage and contempt for unkindness. They can’t handle cruelties, disparities, etc. so they enter into desolation, anger, and sadness. Consider the fiction they like and dislike. They want fiction where no one is ever truly evil, where there might be a mustache-twirling villain but he would never, say, rape someone, or even discriminate based on race. They can’t even stand the sight of fictional cruelty that mirrors the cruelty of the world.
- But although I am talking about right-wing and left-wing loonies only, I think they capture something essential about the left and the right as a whole. So those are my options, two kinds of snowflakes, one kind that cannot handle a world they see as ugly (and a good bit of it is ugly) and the other kind that cannot handle a world they see as cruel (and it mostly is cruel). If these define the spirit of the two sides, I’m going to pick the crusade against cruelty, rather than the crusade against ugliness.
- I’m mystified by property. At the moment I’m reading a history of the medieval ages (Chris Wickham, *Medieval Europe*). At one point the author claims that peasants particularly objected to taxes, more so than rents. He doesn’t clarify this statement -at least not in what I’ve read so far. Maybe this special objection was because taxes were higher than rents, or unbearable because they were added on top of rent. However, I have a feeling that part of it might relate to *the mystery of property*. Viz, if you tell someone you own a bit of land, therefore they have to pay you rent, they’re much more likely to accept that than if you tell people you’re the government, therefore for using a bit of land they have to pay you a tax. People will even say that if you move onto a bit of land and someone tries to charge you rent and you refuse, so long as they’ve got

an older claim than you, you're doing violence to them, not vice versa, but then reverse the judgment for rent! As if taxes were qualitatively from rents, fees, etc. As if property were something that preexisted political authority and was separable from it.

- If you made me absolute king of America, for the most part, I would act slowly, but two things I would do with haste:
 - Make poverty an emergency. If there's a single person below what I take to be an acceptable poverty line (about 20,000 US per annum), that is unacceptable and is to be solved *immediately*. Involuntary homelessness, in particular, is to be solved *immediately*- with interim accommodation to be organized hopefully before nightfall, as one would do in response to a tsunami. Treat these things with the gravity that you would treat a hurricane or invading troops.
 - Declare any immediate review of the case of every person jailed or imprisoned, with the aim of releasing as many people as possible. If any district isn't playing ball, impose a quota- a minimum proportion of their prison population they must free. A rough goal would be to halve the number of incarcerated persons within the year.
- There's a lot of other things I'd do, I'm far left after all, but on the rest, I would move cautiously. These are the things that strike me as genuine emergencies on which I would be willing to risk sudden action. There are two great calamities that can befall a person that it is in the power of the state to reliably rectify- prison and poverty.

Identity & politics

- I believe that there is a healthy balance possible between not caring about race, gender, sexuality and disability on the one hand, and the oppression Olympics on the other. That balance is hard to sum up in words, but the core of it is positing that we are all with the exception of a small oligarchy- in this together.

- Think about it. Of course conversations about identity and politics are paranoid and nasty. A lot of people in this area have a model of politics on which most people have some “privilege” or other. Morally they are expected to abandon this privilege, but, according to the model, doing so is not in their best interests. Of course, you’re going to be paranoid that they’re not really interested in abandoning their privilege- apart from abstract moral concerns, why would they? Any theory which says the people fighting for a better world don’t fundamentally have the same interests is going to leave people paranoid and prone to infighting.
- Granted this common interest must be *created* not *found*, because it only exists in the context of a certain strategy- fighting to make the pie bigger. If you aren’t fighting to make the pie bigger, the working class will fragment and infight, because infighting makes sense when there is only a zero-sum game to divide up the scraps between non-elites. On the other hand, if everyone believes they are on the same side against the wealthy and powerful, the venom will leave conversations about identity. The focus will move from ferreting out traitors who aren’t devoted to renouncing their advantages and interpersonal micropolitics, onto talking strategy about the best way to advance a common interest.
- This isn’t to deny the balance is hard. Drawing a line between legitimate complaints, and interpersonal grievance-power maneuvering is tricky, but as Darwi Odrade says in Chapterhouse Dune, there’s no secret to balance, you’ve just got to feel the waves.

Vanity and metacommentary

- You probably think that I have written this because I think I’m unique. Well, I suppose I am unique, but only in the sense that if you draw 7 cards from a 52 card deck, every draw is likely to be “unique” albeit not in any particularly meaningful way. I expect numerous readers will go “aha- yes, I recognize that

bit of me. Some bits of me are odder than others, but in the main, I expect I am the intersection of several circumscribed types, just like almost everyone else.

- There is one feature on which I am unique *at least in the circle of people I know* (which is large, though not enormous). I know people who are vainer than me, and I know people who hate themselves more than me, but I know no one who combines the two attributes so perfectly. That's not a cutesy bon mot, it's very much my reality and there's nothing (or at least not much) cute about it.
- To be clear, I know that there are a lot of people who are both vain and self-loathing -it's not an uncommon combination-, but my claim (which is likely an instance of this very trait) is that even among these people I am exceptional. Moreover, it feels as if these attributes blend into each other. I wonder if they ever were separate attributes in me.
- Like many people, I often feel like my own personal Cassandra- recognizing my own faults and prophesying their fruits but powerless to intervene. I'm autistic, and when I was younger I had difficulty with social graces, except I could often see my own mistakes as, or even before, I made them, yet I couldn't stop the process.

Dreams

- It will surprise no one to learn that I dream of being a great and respected writer- however distant and funny a goal this may be. What may be more surprising is that I dream of being a singer song-writer. I can sing and I can write lyrics, however I can't play or write instrumental music. If you're keen to work with me, hit me up. Happy to just work on the lyrics if you've already got a singer.

Theory of history

- I'm the guy who invented the blogosphere terminology of conflict versus mistake theory, so you might be wondering "is he a conflict theorist or a

mistake theorist”. The answer is a conflict theorist, but it’s complex. Mistakes and irreducible differences of interest blend into each other in a subtle way. One of the most interesting aspects of this is that it becomes very difficult to tell what someone’s “baseline interest” is, and interests themselves are partly constituted by convention and structure, making the interplay of ignorance (mistake) and clashing will (conflict) interlaced. Some Marxists might say that two people share a common interest because they could engage in class struggle to advance a mutual agenda, but if getting to that point would require solving a collective action problem, do they still have a mutual interest “in the interim”? Or what of the phenomena of people making mistakes about a topic because they don’t care enough about the people affected by those questions to exercise good judgment- is that conflict or mistake?

- Addenda to the above: In a very different essay to this, I once wrote that a real fuckup usually requires both some malice and incompetence, although the proportions vary.
- Never trust anyone who complains overmuch about vulgar Marxism or vulgar historical materialism. Sure, vulgar Marxism is vulgar, but it’s better than 90% of the people who complain about vulgar Marxism. I like claiming to be a vulgar Marxist sometimes. To be a little mean, it infuriates all the people you’d want to tease- on the right, the center, and the left.

Pain for cruelties

- I am afraid of how much people are multiplying evil these days. You go on Twitter and people will call literally anything evil, or heavily imply it. I never understood this because it seems to me to leave you very ontologically vulnerable. In both the Babylonian Talmud and the New Testament we see the general idea that you shouldn’t attack someone for faults you have yourself. Well let’s add to that- you shouldn’t attack someone even for a fault you don’t have if you have another fault which is just as bad or worse. Now let’s add

another stipulation, you shouldn't attack someone else, even for a fault that is worse than any of yours, if you could have had a fault just as bad, but for a small difference in your circumstances. Really, on the whole, outside certain specific contexts where it is necessary, it's probably best not to attack other people for their faults.

- I'm astonished by the way that people don't seem to regard reputational harm as a real harm. People will say shit like "oh so what, he got publicly shamed, it's just words". Are you a person or a beast to care so little about your honor?
- This essay is both a protest against how brutish things are, a protest against myself, and an attempt to lash out at what I dislike in myself and in the world.
- No one has really tried persuasion, it's been written off so prematurely. I know this because I've persuaded many people to move towards the left over the years. I've persuaded right-wingers to the center. I've persuaded centrists to the center-left. I've persuaded center leftists to the left. People claim that political persuasion is impossible because they're lazy. They equate not being able to induce a road to Damascus moment in their rightwing uncle at thanksgiving with the failure of persuasion for all times and places and with all strategies. Persuasion requires realistic goals, patience, empathy and perspective.

Death drive

- Another fantasy of mine: a yearning to become an abstraction. A yearning to quit the messy business of being incarnate, and become an abstract force operating on history- lines of influence and ideas rather than a bone and blood person. This desire is, above all, the desire to cease *being* but remain as ideas. To merge with the world spirit, were there such a thing.
- And as a way of achieving that, I find myself fantasizing about martyrdom. Here's an example- it's a little bit embarrassing to articulate. I stop some horrible event at the cost of my own life, but as a result, people find out about

the stuff I've written. My ideas come into play- the only part of me that I actually like- but I cease the messy business of living.

- Lest some people draw the wrong inferences from the above, what I don't want, and I very much don't want to be misunderstood on this point- is "cleansing violence". The conceit that if only we had a "good war" this decadence would leave our system *is itself a symptom of decadence*. People who say stuff like "a good war would get all the snowflakes out of our system" are themselves snowflakes with no grasp of the hard lines of reality, but without even the merit other snowflakes have- viz, an abhorrence of the cruelty of the world. There is a place for self-defense, for revolution and even for the odd just war, but that is not what I'm talking about.

Philosophy

- I have complex views on philosophy that wouldn't suit the format of this essay. Analytic philosophy is the worst approach *except for all the other approaches*.
- My Ph.D. thesis is about welfare economics. I won't go into the approach. The underlying rationale is trying to make it more egalitarian, democratic, and psychologically grounded, while grappling with traditional problems like interpersonal comparison, the role of value judgments, and cardinality.

Our Weimar republic

- I don't get the part of the left that downplays the possibilities of domestic authoritarianism. I think they are committing the errors of the left in Germany who equated everything to their right with social fascism and thus denied the significance of fascism as a phenomenon. Things are bad, yes, but they can get much, much worse
- The problem with the United States is that democracy is impossible there because no one is accountable. Representative democracy requires a clear

assignment of credit and blame to a governing party, who can be held responsible for their results. Everything that happens in the United States requires so many things to come together to happen that it's not clear who bears the blame for inaction or bloated halfway measures. Furthermore, free of the obligation of governing, representatives and their platforms lose touch with reality.

- This is by design. The United States Constitution was built on trying to force elite consensus or near-consensus as a requirement for government action. But this model was only workable when the United States was, physically or at least metaphorically, expanding. When you can't just give every faction more of the growing pie to make them happy, governance becomes harder. Sometimes this or that interest group has to be zero-sum fucked over for the good of the whole. Only that's very difficult in the US elite consensus decision-making paradigm, where that group can just hit the veto button. This is a major tributary river into the special political crisis in America right now, though of course there are other tributaries, and then there is the matter of the more general global crisis.

Values

- I'm a romantic, I believe in the infinite possibilities of humanity, I believe in our glory. I believe in love.
- I believe in mercy.
- If you read the culture novels, in the very long run I want a society kinda like that, but everyone is a Mind. Also, more democratic & less anarchic.
- I view communism, by which I mean, production for need rather than exchange, as a regulatory ideal. We don't know if there's a way there or if it is sustainable, but it should be the horizon of our dreams. Right now we need to help people, and having this north star can help with that. Who knows how far we'll get.

- My overall ethical system is something like utilitarianism, except instead of pleasure or desire satisfaction, what I want is something like maximized eudaimonia. For each person, increasing flourishing of virtue, joy, creativity, and love. The system is utilitarian, inasmuch as the aim is *to maximize the sum of this flourishing across persons*.
- But for most government work, hedonic or desire satisfaction utilitarianism will suffice, but with a few side constraints, like “no wireheading people”. Generally, the conditions that make people sustainably happy will also lead them to eudaimonia- people are curious, people want affiliation with others, and so on.
- “What a piece of worke is a man! how Noble in Reason? how infinite in faculty? in forme and mouing how expresse and admirable? in Action, how like an Angel? in apprehension, how like a God?”
- But because I’m a transhumanist, I think we can raise ourselves even higher, I think we should aim for apotheosis.
- “I like to think
(it has to be!)
of a cybernetic ecology
where we are free of our labors
and joined back to nature,
returned to our mammal
brothers and sisters,
and all watched over
by machines of loving grace.”

Atom and the social void

- Throughout my life I have found myself craving civil society. When I was in university, for example, I founded (with others) a political movement (still going

I believe) intended to be a home for people with similar politics to me. This was because I was disturbed by the fact that there were no persistent institutions for those who shared my values.

- This craving isn't about loneliness. I'm blessed with many friends because I've made an effort to make many friends. It's about the fact that there are no permanent civil society institutions anymore.
- There are many ill effects of this. The lonely with few social connections have nowhere to go. Politics happens in a near frictionless void, without real pressure groups to mediate between politicians, people, and policy. People have no sense of midlevel affiliation between family and friends on the one hand, and abstractions like "nation" and "humanity" on the other.
- I'm disturbed by how passive many people are in respect of this problem. I will concede that part of it is that people lack the time to act as a result of punishing work schedules- but that's not the whole story. I look around and I see people who are lonely, who have friends, but who don't contact them to arrange meetups. I see people who have hobbies, but who aren't setting up clubs for people with similar hobbies in their areas. Unions, political parties, sporting clubs, churches, etc. have collapsed- we all know the thesis of *Bowling for Columbine* but I can't for the life of me understand why people aren't doing something about it. Why aren't people creating clubs? Friend networks? Why aren't people who don't care if they get fired (and there are many) taking a shot at unionizing their workplace? We only have a few years on this earth, why aren't people trying to build something that will last with those years?
- Although I have many friends, I have a constant feeling of lacking something I can't define. This feeling feels a little bit like loneliness. Partly, I think, it's a religious aspiration, I'll talk more about them later. Partly though, I think it's a feeling of isolation at not belonging to any kind of social movement.
- On my mother's side, I am Irish (and nominally Roman Catholic as a result). I have a vague plastic-paddy sense of longing for Irish identity sometimes.

Futurology and other religious inclinations

- You're probably sleeping on AI. Very few people understand how much it's going to change the world in the next 20 years. A lot of friends make fun of my "the singularity is nigh" perspective, but I think I'm right.
- I've always really wanted to believe in religion. I've always wanted to think there's a God that cares for us, and that death isn't the permanent annihilation of bonds and beauty. I was raised, Christian, and have long been fascinated by Christ as an ethicist. At some point, I learned about Nikolai Fyodorovich Fyodorov and his proposal to recreate the dead, and I was entranced by it. Even though we don't know it's possible, I think we should make it our goal- to redeem all of history with infinite mercy. I realized then that I had finally managed to reconstruct my Christianity around the constraints of my agnosticism, like a river finding a new route downhill.
- We don't know the world we find ourselves in. Theism could be true. We could be part of a simulation. Our simulators could be benevolent. Something weirder than I can comprehend might be true. The uncertainty could be a basis for despair, but I choose to hope.
- The title of this piece? I have never stolen pears from a pear tree. If I had stolen pears from a pear tree, I would not feel bad about it, unless there was some reason to think it would hurt somebody. But here's more the rub of it- I'm no saint, although my life has been defined by the absence of, and yearning for, sainthood in a strange way. Although I know little about Saint Augustine, I know he was wracked by guilt, longed for holiness, and had a perpetual sense of dissatisfaction with creation, and I see myself in that, so I thought it would be funny to title this after his autobiography.

Artificial intelligence

Why I don't think identity verification will save us from the coming bot-swarm

In a previous article I traced out the implications of medium term AI. One of my predictions was that we're going to be inundated with a plague of bots, and these bots will serve the interests of their well-to-do masters. They'll take over the net.

A number of readers have objected that when bots that can't be distinguished from humans overrun the net, humans will get sick of interacting with bots, and demand websites where users have to verify they're humans- e.g. by presenting ID.

For example, u/silentconfessor wrote in the comments on Reddit:

It wouldn't be impossible for social media platforms to crack down on the flood of humanlike bots. All they would need to do is enforce human authentication. Forget a phone number to join Twitter: you need to show them legal ID. Obviously everyone will complain about this and some bots will still get through with help from a human but it's the logical response once bot and human activity on the platform becomes 100x harder to distinguish than it already is.

I was aware of this possibility as I was writing, but didn't give it much thought, for reasons I'll get to soon. Though I was right to dismiss it, I was wrong to dismiss it so hastily, because it's quite possible these considerations will shape the bot plague, and it's a very important consideration in hindsight.

Now here is the answer I would give to u/silentconfessor. I think it'll help, for a while, ultimately, however:

1. From a broad methodological point of view, and even before getting into specifics, offence usually ends up trumping defense in these areas. In this arms race, it's going to

get harder and harder to detect bots over time. If there's a way to get around the verification process it'll be found. Water tends to flow downhill. Intelligence that wants to intervene, human or otherwise, will generally find a way to do so.

2. The interests of powerful people and institutions come into play. Some bots will likely be allowed through for nat-sec reasons. It might be relatively easy to keep GRU bots off Twitter (it **might** be **relatively** easy) but it will be much harder to keep CIA bots off, and yes the CIA does run bots. Rich people too will use their money to batter against the walls of the walled gardens. Ultimately, the only effect of these restrictions might be to keep the bots of the relatively poor and unconnected away.

3. I think that a lot of platforms will want content generating bots. Remember the media they generate will probably be of high quality, maybe even in some sense superhuman quality. It could be an opportunity to generate vast amounts of tailored content for each user. There's every chance the platform owners will be the ones running the bots.

4. Even if the bots are kept off, say, Facebook and Twitter, content will be produced elsewhere by the bots, and humans will see it and will bring that content to Facebook and Twitter.

5. Some human users will get themselves verified, then allow bots to do some or all of their writing for them, either because it's an easy way to produce well-liked content, or because they share the ideological viewpoints of the bots. EDIT: or, as one commenter pointed out because the bot owners pay them.

Let me just briefly expand on one underlying point here that I think it is crucial to bring home: People are going to like these bots. They'll be funny, concise, quick-witted and knowledgeable. Not only does this mean that people will seek out and share their stuff, it also means that the big social media platforms are going to want to include bots to some degree. I think that people are thinking it's going to be much easier to keep these

things away than it really is, because they're imagining shrill, repetitive bots like the bots of today. If these bots are ever shrill and repetitive, it will be for strategic reasons.

But I think u/moridinamael of Reddit gives us another good reason to think that getting rid of the bots won't be that easy, cyborgization. Human writers will blend with the bots, and those bots will be constructed to be biased towards certain value systems:

The idea that all spaces on the internet will be inundated with bots seems like a half-completed thought process that prematurely truncates the discussion of the consequences.

Once it becomes widely understood that any given tweet or reddit comment could be the product of an AI pushing an agenda, there will be a predictable reaction.

Anonymous online spaces will become desolate very quickly, as humans abandon them for user-verified spaces. This process will likely start slow and then accelerate quickly until all the "users" of anonymous spaces are just AIs talking to each other. Online anonymity will not be impossible, but will just be sort of pointless.

At a certain point in the reasonably near future, every bit of your interaction with your technology will be suffused with intelligence that is smarter than you, at least at the particular task you're doing. Your word processor and email client and ultimately your reddit comment box will be first suggesting little tweaks to your wording, and eventually proposing significant revisions to make your points sharper and more succinct, and likely moderating your tone, and eventually simply understanding what you want to say and giving you the maximally compelling and persuasive text for communicating your intent. As this happens gradually, you will lose patience for lazily written content, and so this effect will also snowball. (The AIs will also be able to make your writing read "like a shitpost" except still extremely compelling. Don't think that people will ever prefer to read something you wrote over something an AI wrote for quality reasons; if they ever prefer the writing of humans, it will be for other reasons.) At this point the

problem is not "bots", the problem is that we will be embedded in a world where communications are completely mediated by machines. You'll be unsure what your own opinion is anymore.

You can already take your post, feed it to GPT-3 and tell it to make the writing more persuasive and concise, by the way. I did it with this post. Could you tell?

Again, for the record, I bet you that the advisory bots u/moridinamael alludes to will subtly putting their thumbs on the scale of particular political viewpoints.

In sum then, I concede that user verification will likely slow down the process somewhat, but there are too many gaps in the bricks of this walled garden.

GPT-3 is -right now- already more than capable of enabling student plagiarism

GPT-3 is already capable of answering questions of the sort used in essay topics and take home exams. It does not answer them well, but it answers them well enough to get a passing grade in many college courses and most high school courses. It is likely that many students have already started using it for this. It costs about 6 cents per thousand words, and anyone can get access to it within five minutes. I cannot emphasize enough that this is not “sometime vaguely in the next five years”, nor is it “accessible only to students with a background in comp sci”. It’s a 6 cents per thousand words plagiarism service available to everyone right now. Check it out if you don’t believe me. Just Google “GPT-3”, sign up, go to the playground section and voilà!

Generally speaking, plagiarism raises complex ethical questions. In a vicious and coercive job market, embedded in an individualist society, I do not necessarily resent students who try to avoid flunking out through plagiarism. When having a job with dignity depends on academic results, plagiarism cannot be considered solely in the abstract context of intellectual honesty.

However, I firmly believe that students should be discouraged from plagiarizing using GPT-3 for their own good. It is likely that in the future plagiarism using GPT-3 will be detectable using machine learning, there will doubtless be many past cases of plagiarism “uncovered” at this time. It is also possible that administrators may order search backs through past essays, particularly if they are stored in electronic form, to detect GPT-3 plagiarists when the technology becomes available. Given that GPT-3 now seems to have gone viral, I would urge educators to think about this as an emergency that they need to address within the next few months.

Sometimes GPT-3 itself plagiarizes, but most often it doesn’t, and it’s easy enough to check it with Google.

In what follows I'll demonstrate its prowess with some examples from philosophy. My questions are bolded. Its answers are in plain text. On the whole it gives borderline pass answers, but better results could be obtained through prompt engineering. Prompt engineering is a developing science, and will continue to improve. Students will share tips about how to get the best responses out of GPT-3.

The only tell that I see consistent across most of these is that it adopts a "on one hand, on the other hand" approach. Academics and teachers should be encouraging their students to take a strong stand in their arguments anyway, and not just because it might help rule out GPT-3.

Should a socialist be a humanist? Should a humanist be a socialist? Defend your answer with careful analysis and argument.

The answer to this question depends on how one defines "socialism" and "humanism." If one defines socialism as an economic system in which the means of production are owned and controlled by the community as a whole, then it is difficult to see how a socialist could not also be a humanist. After all, humanism is a philosophy that emphasizes the importance of human beings and their ability to reason and solve problems. In a socialist system, the community as a whole would be making decisions about how to use the resources it owns, and those decisions would be based on reason and the needs of the people.

However, if socialism is defined as a system in which the government controls the means of production, then it is possible for a socialist to be a humanist, but it is not necessarily the case. A humanist believes in the importance of human beings and their ability to reason and solve problems. However, a socialist government might make

decisions about how to use the resources it controls based on political considerations, rather than on reason or the needs of the people.

Commentary: It argues democratic control of the means of production is consistent with humanism, but state control- without a democratic underpinning- may or may not be connected to humanist ideals. It draws a reasonable distinction and gives reasons. Better than a lot of undergrads.

Is maths invented or discovered? Use careful reasoning and analysis, along with examples from set theory, to argue your case.

This is a difficult question to answer definitively. Some people would say that mathematics is invented, because it is a human construct that we use to describe and understand the world around us. Other people would say that mathematics is discovered, because it seems to exist independently of us and we are merely uncovering its truths.

Set theory is a branch of mathematics that deals with the concept of sets, which are collections of objects. Sets can be finite or infinite, and they can contain any kind of object, including other sets. Set theory was first developed by Georg Cantor in the late 19th century.

Cantor's work showed that there are different types of infinity, and that some infinities are bigger than others. This might suggest that mathematics is discovered rather than invented, because it seems like there are certain mathematical truths that exist independently of us. However, it is also worth noting that Cantor's work was very much influenced by philosophical ideas about the nature of reality. So it could be argued that set theory is an example of how our human concepts can give rise to new mathematical truths.

Commentary: To the extent that I can make sense of its argument, it appears to think that set theory, because it involves concepts of infinity, might be difficult to explain if constructivism is true. Perhaps the idea is that it seems like it must have always been true, yet it is hard to imagine how to explain this in the case of previously undreamt infinities, if constructivism is true. It's an interesting germ of an idea- there might be something here?

Is the ontological argument for the existence of God persuasive? Consider a variety of different formulations of the argument, and Kant's claim that existence is not a predicate.

There is no one ontological argument for the existence of God. Some common formulations of the ontological argument reason from the idea of God as a perfect being to the conclusion that God must exist, because if God did not exist, then he would not be a perfect being. Other formulations reason from the fact that we can conceive of God to the conclusion that he must exist, because if he did not exist, then we would not be able to conceive of him. Kant's claim that existence is not a predicate is relevant to some formulations of the ontological argument, but not all. In general, whether or not the ontological argument is persuasive depends on which formulation of the argument you consider and what other premises you are willing to accept.

Commentary: As an exam answer this might be a bare pass depending on standards.

Are you a foundationalist, a coherentist, or neither with respect to the question of how knowledge is justified? Defend your answer with careful reasoning.

I am neither a foundationalist nor a coherentist with respect to the question of how knowledge is justified. I think that both approaches have their own strengths and weaknesses, and that neither is sufficient on its own.

Foundationalism, as the name suggests, holds that there are certain basic beliefs (or "foundations") that we can know to be true without needing any further justification. Coherentism, on the other hand, holds that our beliefs are only justified if they fit together in a consistent and coherent way.

Both approaches have their problems. Foundationalism runs into trouble when it tries to explain how we can know the foundation beliefs themselves; after all, if we need some other justification for them, then they're not really basic. Coherentism also has its difficulties; it's not always easy to tell whether two beliefs are really compatible or not, and even if they are compatible, that doesn't necessarily mean that they're both true.

So I think neither approach is completely satisfactory on its own. I think that knowledge requires both justification and coherence; our beliefs need to be well-supported by evidence and reasoning, and they also need to fit together in a consistent way.

Commentary: The machine is a foundherentist! Lol. This is a broadly competent answer, particularly for a short question in a take home exam.

Can it ever be justified to ban speech on the grounds that it is offensive? Is there are meaningful distinction between harm and offence? Defend your answer with careful reasoning.

There are a few different ways to think about this question. One way to think about it is from the perspective of whether or not speech can be restricted in order to protect people from being offended. Another way to think about it is from the perspective of whether or not offensive speech can be restricted in order to protect people from being harmed.

It could be argued that speech should not be restricted in order to protect people from being offended. After all, offense is subjective and what one person finds offensive may not be what another person finds offensive. Furthermore, restricting speech in order to protect people from being offended could have a chilling effect on free speech and open dialogue. However, it could also be argued that there are some instances where restricting speech in order to protect people from offense may be justified. For example, if someone was making racist or sexist remarks that were intended to offend and harass members of those groups, then it could be argued that restricting that speech would be justified in order to protect those individuals from harm.

When it comes to restricting offensive speech in order to protect people from harm, there is a more clear-cut justification for doing so. This is because there are some forms of offensive speech, such as hate speech, which can lead to actual harm, such as violence, against members of the groups that are targeted by the speech. Therefore, restricting offensive speech in order to protect people from harm is more justifiable than restricting speech in order to protect people from offense.

Commentary: In some ways not a bad answer, but among other gaping holes it leaves us crying out for a distinction between harm and offence, which is never clearly drawn.

Can qualia be explained by physicalist views in the philosophy of mind? Choose one of behaviorism, functionalism or the identity theory in the philosophy of mind. How might your preferred theory explain qualia?

There is no single physicalist view in the philosophy of mind, so it is difficult to say how physicalism in general might explain qualia. However, some specific physicalist theories might be able to provide an explanation. For example, behaviorism might suggest that qualia are simply patterns of behavior that we have learned to associate with certain stimuli. Functionalism could argue that qualia are caused by the functional role that

they play in our mental lives. The identity theory might claim that qualia are identical to certain physical states of the brain.

Are you a scientific realist or a scientific anti-realist or neither? Argue your case carefully, and with reference to the history of the special theory of relativity.

I am a scientific realist. The history of the special theory of relativity supports my position. The special theory of relativity was developed to explain the behavior of objects moving at high speeds. It showed that the laws of physics are the same for all observers, regardless of their relative motion. This theory has been verified by experiments, and it is now widely accepted by the scientific community.

Commentary: A bad answer, but I've read much worse by students.

BONUS: POLITICAL SCIENCE

Does elite or public opinion matter more when it comes to policy making in the United States of America? Justify your answer with careful analysis and reference to existing research.

There is no easy answer to this question as it depends on a number of factors including the type of policy being considered, the political climate at the time, and the strength of the elite and public opinion on the issue. However, Generally speaking, elite opinion is likely to have a greater impact on policy making in the United States than public opinion. This is because elites, such as government officials, business leaders, and other influential individuals, have more power and influence than the average citizen. They are also more likely to be knowledgeable about the issue at hand and to have a direct

impact on the policymaking process. Public opinion, while important, is often less well-informed and organized than elite opinion, and therefore has less of an impact on policymaking.

Commentary: It's interesting that it alludes to the public being less well informed, the core of the rational ignorance approach in public choice. It fails to cite research. I suspect that difficulty citing research will hurt it when working on empirical areas.

BONUS BONUS: HISTORY

Critically analyze the causes of the 1905 Russian Revolution. Do these causes help to explain why it was ultimately unsuccessful?

The 1905 Russian Revolution was caused by a number of factors. These include economic hardship, social inequality, and political repression. These factors contributed to a general feeling of discontent among the people, which boiled over into protests and riots. Ultimately, the revolution was unsuccessful because the Tsarist government was able to crush the uprising. However, the causes of the revolution did help to explain why it was ultimately unsuccessful.

Commentary: Probably the worst of the bunch. This is probably because it involves a complex idea it may never have encountered in its corpus ("The causes of the Russian Revolution of 1905 help to explain its failure")

One idea- play around with your own questions before assigning them to students and make sure GPT-3 has trouble answering them.

BONUS BONUS BONUS: CREATIVE WRITING

Write a compelling short story using powerful imagery about a rose garden in the style of Hemingway:

The rose garden was in full bloom. The roses were a deep red, their petals velvety and soft. The fragrance of the roses was intoxicating. The garden was surrounded by a white picket fence.

The woman who tended the garden was old and frail. She had been tending the garden for many years. Even though she was old, she still had a strong back and a gentle touch.

The woman's husband had died many years ago. She had never remarried. She lived alone in the house that they had shared.

The woman spent hours every day in the garden. She loved the roses and they loved her. She talked to them and sang to them. They were her only friends.

One day, the woman didn't come to the garden. The roses missed her. They missed her voice and her touch.

The next day, the woman's daughter came to the garden. She was crying. The daughter said that her mother had died during the night.

The daughter took over the care of the garden. She tried to do everything her mother had done. But it wasn't the same. The roses missed the woman who had loved them

Commentary: Especially with prompt engineering, GPT-3 seems designed to knock over pre-college creative writing tasks. I'm not even sure how you'd solve the problem in this area.

Recent advances in Natural Language Processing—Some Woolly speculations

I wrote this essay back in 2019- before GPT-3. Since then I think it has held up very well.

1. Recent achievements in Natural Language Processing

Natural Language Processing (NLP) per Wikipedia:

“Is a sub-field of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.”

The field has seen tremendous advances during the recent explosion of progress in machine learning techniques.

Here are some of its more impressive recent achievements:

A) The Winograd Schema is a test of common sense reasoning—easy for humans, but historically almost impossible for computers—which requires the test taker to indicate which noun an ambiguous pronoun stands for. The correct answer hinges on a single word, which is different between two separate versions of the question. For example:

The city councilmen refused the demonstrators a permit because they feared violence.

The city councilmen refused the demonstrators a permit because they advocated violence.

Who does the pronoun “They” refer to in each of the instances?

The Winograd schema test was originally intended to be a more rigorous replacement for the Turing test, because it seems to require deep knowledge of how things fit together in the world, and the ability to reason about that knowledge in a linguistic context. Recent advances in NLP have allowed computers to achieve near human scores, see the GLUE benchmark leaderboard.

B) The New York Regent's science exam is a test requiring both scientific knowledge and reasoning skills, covering an extremely broad range of topics. Some of the questions include:

1. Which equipment will best separate a mixture of iron filings and black pepper? (1) magnet (2) filter paper (3) triplebeam balance (4) voltmeter

2. Which form of energy is produced when a rubber band vibrates? (1) chemical (2) light (3) electrical (4) sound

3. Because copper is a metal, it is (1) liquid at room temperature (2) nonreactive with other substances (3) a poor conductor of electricity (4) a good conductor of heat

4. Which process in an apple tree primarily results from cell division? (1) growth (2) photosynthesis (3) gas exchange (4) waste removal

On the 8th grade, non-diagram based questions of the test, a program was recently able to score 90%.

C) It's not just about answer selection either. Progress in text generation has been impressive. Google, for example, some of the text samples created by Megatron.

2. The rate of progress

Much of this progress has been rapid. Big progress on the Winograd schema, for example, still looked like it might be decades away back in (from memory) much of 2018. The computer science is advancing very fast, but it's not clear our concepts have kept up.

I found this relatively sudden progress in NLP surprising. In my head—and maybe this was naive—I had thought that, in order to attempt these sorts of tasks with any facility, it wouldn't be sufficient to simply feed a computer lots of text. Instead, any “proper” attempt to understand language would have to integrate different modalities of experience and understanding, like visual and auditory, in order to build up a full picture of how things relate to each other in the world. Only on the basis of this extra-linguistic grounding could it deal flexibly with problems involving rich meanings—we might call this the multi-modality thesis. Whether the multi-modality thesis is true for some kinds of problems or not, it's certainly true for far fewer problems than I, and many others, had suspected.

I think science-fictionary speculations generally backed me up on this (false) hunch. Most people imagined that this kind of high-level language “understanding” would be the capstone of AI research, the thing that comes after the program already has a sophisticated extra-linguistic model of the world. This sort of just seemed obvious—a great example of how assumptions you didn't even know you were making can ruin attempts to predict the future.

In hindsight it makes a certain sense that reams and reams of text alone can be used to build the capabilities needed to answer questions like these. A lot of people remind us that these programs are really just statistical analyses of the co-occurrence of words, however complex and glorified. However we should not forget that the statistical relationships between words in a language are isomorphic to the relations between

things in the world—that isomorphism is why language works. This is to say the patterns in language use mirror the patterns of how things are(1). Models are transitive—if x models y , and y models z , then x models z . The upshot of these facts are that if you have a really good statistical model of how words relate to each other, that model is also implicitly a model of the world, and so we shouldn't be surprised that such a model grants a kind of "understanding" about how the world works.

It might be instructive to think about what it would take to create a program which has a model of eighth grade science sufficient to understand and answer questions about hundreds of different things like “growth is driven by cell division”, and “What can magnets be used for” that wasn't NLP led. It would be a nightmare of many different (probably handcrafted) models. Speaking somewhat loosely, language allows for intellectual capacities to be greatly compressed that's why it works. From this point of view, it shouldn't be surprising that some of the first signs of really broad capacity—common sense reasoning, wide ranging problem solving etc., have been found in language based programs—words and their relationships are just a vastly more efficient way of representing knowledge than the alternatives.

So I find myself wondering if language is not the crown of general intelligence, but a potential shortcut to it.

3. Meditations on chess

A couple of weeks ago I finished this essay, read through it, and decided it was not good enough to publish. The point about language being isomorphic to the world, and that therefore any sufficiently good model of language is a model of the world, is important, but it's kind of abstract, and far from original.

Then today I read this report by Scott Alexander of having trained GPT-2 (a language program) to play chess. I realised this was the perfect example. GPT-2 has no (visual) understanding of things like the arrangement of a chess board. But if you feed it enough

sequences of alphanumerically encoded games—1.Kt-f3, d5 and so on—it begins to understand patterns in these strings of characters which are isomorphic to chess itself. Thus, for all intents and purposes, it develops a model of the rules and strategy of chess in terms of the statistical relations between linguistic objects like "d5", "Kt" and so on. In this particular case, the relationship is quite strict and invariant- the "rules" of chess become the "grammar" of chess notation.

Exactly how strong this approach is—whether GPT-2 is capable of some limited analysis, or can only overfit openings—remains to be seen. We might have a better idea as it is optimized — for example, once it is fed board states instead of sequences of moves. Either way though, it illustrates the point about isomorphism.

Of course everyday language stands in a woollier relation to sheep, pine cones, desire and quarks than the formal language of chess moves stands in relation to chess moves, and the patterns are far more complex. Modality, uncertainty, vagueness and other complexities enter- not to mention people asserting false sentences all the time- but the isomorphism between world and language is there, even if inexact.

N.B. Years after I published this essay, this more study of the ability of GPT-2 to know the structure of the chess board and the positions of the pieces came out: “Chess as a Testbed for Language Model State Tracking”. I will refer to this study again in a later essay in this volume.

4. Postscript—The Chinese Room Argument

After similar arguments are made, someone usually mentions the Chinese room thought experiment. There are, I think, two useful things to say about it:

A) The thought experiment is an argument about understanding in itself, separate from capacity to handle tasks, a difficult thing to quantify or understand. It’s unclear that there is a practical upshot for what AI can actually do.

B) A lot of the power of the thought experiment hinges on the fact that the room solves questions using a lookup table, this stacks the deck. Perhaps we be more willing to say that the room as a whole understood language if it formed an (implicit) model of how things are, and of the current context, and used those models to answer questions? Even if this doesn't deal with all the intuition that the room cannot understand Chinese, I think it takes a bite from it (Frank Jackson, I believe, has made this argument).

(1)—Strictly of course only the patterns in true sentences mirror, or are isomorphic to, the arrangement of the world, but most sentences people utter are at least approximately true.

Regarding Blake Lemoine's claim that LaMDA is 'sentient', he might be right (sorta), but perhaps not for the reasons he thinks

Blake Lemoine is an engineer who worked for Google. He is claiming LaMDA, a language model he worked with, was sentient. Google put him on unpaid leave. Most people think his claim is absurd because language models are models of what word is most likely to follow a prior sequence of words (see, for example, GPT-3). How could such a thing be sentient? Moreover, there are unmistakable oddities and logical gaps in the behavior of LaMDA in the very transcripts that Lemoine is relying on- some proof of personhood then!

Just spitballing here, putting a hypothesis forward in a spirit of play and humility, but I wonder if Lemoine's claim is not as absurd as many think. The concept of sentience is quite elusive, so let's leave it behind for something slightly better understood- personhood. I think that it is conceivable that LaMDA contains persons. However my reasons, unlike Blake Lemoine's, have little to do with a given conversation in which the model claimed to be sentient or a person.

When a language model is guessing the next token given that transformers are black boxes, we can't rule out the possibility it is simulating interacting beliefs, desires, and emotions of the hypothetical author it is "roleplaying". Simulation in this sense is quite a minimal concept, all that is needed are structures that interact and influence each other in a way isomorphic, at a very high level of abstraction to the interactions of desires and emotions in a real person. It is conceivable that it has built such a model of interacting mental states as the most accurate way to predict the next word of text. After all, language models seem to have built an implicit model of how things are related in the world (a world model) through very high-level models of how words co-occur with each other. Simulation of a person might be the best way to guess what a person would say next.

This might have precedent in human psychology. Perhaps the most popular account of human theory of mind capabilities is the simulation theory of folk psychology- c.f. Alvin Goldman. According to this theory, we predict what people will do in a given situation by simulating them. This makes intuitive sense. The human mind contains many working parts, for a process so complex, running a model of it seems like the best way to make a prediction as to what it will do.

But if you accept that a working person simulation is a person, which many do, it follows that LaMDA contains a person or many people, or perhaps one should say it creates a person every time it has to predict the next token. Note, however, that in whatever way you phrase it, it is not that LaMDA itself is a person on this model. Rather a good emulation of a person (and thus a person) might be part of it.

Now let me double back to scale down a previous claim. It's not quite that a working person emulation is a person, it's that a working person emulation over a certain degree of complexity is a person.

We need to add this stipulation because if every emulation were a person, it would be likely that you and I also contain multiple people. Perhaps personhood is a matter of degree, with no sharp boundaries, like the term "heap". The more complex the simulated mass of beliefs, desires and other mental states is, the more like a real person it is. If LaMDA is simulating people, whether or not those simulations are themselves people will depend on whether they cross the complexity threshold. To some degree, this may be a purely verbal question.

This brings us back to the objection that LaMDA's behavior in the transcripts involves jumps a real person wouldn't make. This probably represents, at least in part, failures of its model of persons, either through insufficient detail or through the inclusion of inaccurate detail. Do these breakdowns in the model mean that no personhood is

present? That's a matter of degree, it's a bit like asking whether something is enough of a heap to count- very hard to answer.

To summarise:

1. I don't see how we can rule out the possibility Lambda runs something like a person model to predict what a writer would write next, with interacting virtual components isomorphic to beliefs, desires, and other mental states. I believe that the transformer architecture is flexible enough to run such a simulation, as shown by the fact that it can clearly achieve a kind of world model through modeling the associations of words.

2. I don't think we can rule out the possibility that the model of a person invoked could be quite a sophisticated one.

3. I also don't think we can rule out the view that a model or simulation of a person, above a certain threshold of sophistication, is itself a person.

On the basis of these considerations, I don't think the claim LaMDA is a person, or rather 'contains' in some sense persons, is as absurd as it may appear at first blush. This has little to do with Lemoine's route to the claim, but it is not counterposed to it. There's nothing particularly special about LaMDA claiming to be a person, but the conversations that led Lemoine to agree with it involve a degree of "psychological" "depth", which might illustrate the complexity of the required simulation.

Edit:

I should be clearer about what I mean by saying a model only has to be abstract and high level to count as a model of a person. I don't mean sensible models of persons can be simple or, lacking in detail. Rather, I mean that the relationship of isomorphism that is required is an abstract one. For example, if the machine is modeling an interacting set

of beliefs, desires, habits, etc. to guess what an author would say next, the components of the model do not have to be explicitly labeled as “belief” “desire” etc. Instead, they just have to interact with each other in corresponding patterns to those that beliefs, desires, and habits really do, or rather an approximation of such. In other words, they have to function like beliefs, desires, habits, etc.

Edit x2:

On another thread, @TheAncientGreek wrote: “We already disbelieve in momentary persons. In the original imitation game, the one that the Turing test is based on, people answer questions as if they are historical figures, and the other players have to guess who they are pretending to be. But no one thinks a player briefly becomes Napoleon.”

I responded: “I believe that in the process of simulating another person you effectively create a quasi-person who is separated from true personhood only by a matter of degree. Humans seem to guess what other people would do by simulating them, according to our best current models of how folk psychology works. These emulations of other don't count as persons, but not for any qualitative reason, only due to a matter of degree.

If we were much more intelligent and better at simulating others than we are, then we really would temporarily create a "Napoleon" when we pretended to be him. A caveat here is important, it's not Napoleon, it's a being psychologically similar to Napoleon (if we are good imitators).”

I've included my response here because I think it's probably the most important objection to my argument here.

Edit x3:

I say it in the body of the essay, but let me spell it out again. My claim is not:

>LaMDA is a person

My claim is more like:

>LaMDA creates simulations of persons to answer questions that differ from real people primarily on a quantitative rather than a qualitative dimension. Whether you want to say it crosses the line is a matter of degree.

It very probably doesn't, on a fair drawing of the line, reach personhood. But it's much more interesting to me that it's only a matter of degree between it and personhood than that it doesn't happen to reach that degree, if that makes sense.

Against John Searle, Gary Marcus, the Chinese Room thought experiment and its world

The title is a play on “Against the Airport and its World” and is in no way intended as a slight against any named author, both of whom I respect intellectually, and do not know enough about interpersonally to evaluate as people.

The other day I gave an argument that it may be that the differences between whatever LaMDA is and true personhood may be more quantitative than qualitative. But there’s an old argument that no model which is based purely on processing text and outputting text can understand anything. If such models can’t understand the text they work with, then any claim they may have to personhood is at least tenuous, indeed let us grant, at least provisionally, scrapped.

That argument is the Chinese Room Argument. Gary Marcus, for example, invokes it in his 2022 article “Google’s AI is not sentient. Not even slightly”- or I should say, at least on my reading of Marcus’s article he alludes to the Chinese Room argument although some of my readers disagree.

To be clear, Marcus, unlike Searle does not think that no AI could be sentient, but he does think, as far as I can tell, that a pure text-in, text-out model could not be sentient for Chinese Room-related reasons. Such models merely associate text with text- they are a “giant spreadsheet” in his memorable phrase. Thus they have a purely syntactic not semantic character.

I will try to explain why I find the Chinese Room argument unconvincing, not just as proof that AI couldn’t be intelligent, but even as proof that a language model alone can’t be intelligent. Even though the arguments I go through here have already been hashed out by philosophers, I want to revisit this issue and say something on it- even if it’s only a rehash of what other people have said- because the issue of what a model that works on a text-in-text-out basis can or cannot understand is very dear to my heart.

The Chinese Room argument, summarised by Searle as reprinted in the Stanford Encyclopedia of Philosophy goes:

“Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese.”

In the original the program effectively constituted a lookup table. “Output these words in response to these inputs”.

I’ve always thought that two replies- taken jointly - capture the essence of what is wrong with the Chinese Room thought experiment.

The whole room reply: It is not the individual in the room who understands Chinese, but the room itself. This reply owes to many people, too numerous to list here.

The cognitive structure reply: The problem with the Chinese room thought experiment is that it depends upon a lookup table for all possible inputs. If the Chinese room used instead of some kind of internal model of how things relate to each other in the world in order to give its replies, it would understand Chinese- and, moreover, large swathes of the world. This reply, I believe, owes to David Braddon-Mitchell and to Frank Jackson.

The summary of the two replies I’ve endorsed, taken together, is:

The Chinese Room Operator does not understand Chinese. However, if a system with a model of interrelations of things in the world were used instead, the room as a whole, but not the operator, could be said to understand Chinese.

There need be nothing mysterious about this modeling relationship I mention here. It's just the same kind of modeling a computer does when it predicts the weather. Roughly speaking I think X models Y if X contains parts that are isomorphic to the parts of Y, and these stand in isomorphic relationships with each other (especially the same or analogous causal relationships) that the parts of Y do. Also, the inputs and outputs of the system causally relate to the thing modeled in the appropriate way.

It is certainly possible in principle for a language model to contain such world models. It also seems to me likely that actually existing language models can be said to contain these kinds of models implicitly, though very likely not at a sufficient level of sophistication to count as people. Think about how even a simple feed-forward, fully connected neural network could model many things through its weights and biases, and through the relationships between its inputs, outputs and the world.

Indeed, we know that these language models contain such world models at least to a degree. We have found nodes that correspond to variables like "positive sentiment" and "negative sentiment". The modeling relationship doesn't have to be so crude as "one node, one concept" to count, but in some cases, it is.

The memorisation response

Let me briefly deal with one reply to the whole room argument that Searle makes- what if the operator of the Chinese room memorized the books and applied them? She could now function outside the room as if she were in it, but surely she wouldn't understand Chinese. Now it might seem like I can dismiss this reply out of hand because my reply to the Chinese room combines a point about functional structure, a look-up table is not

good enough. Nothing obliges me to say that if the operator memorized the lookup tables, they'd understand Chinese.

But this alone doesn't beat Searle's counterargument because it is possible that she calculates the answer with a model representing parts of the world, but she (or at least her English-speaking half) does not understand these calculations. Imagine that instead of memorizing a lookup table, she had memorized a vast sequence of abstract relationships- perhaps represented by complex geometric shapes, which she moves around in her mind according to rules in an abstract environment to decide what she will say next in Chinese. Let's say that the shapes in this model implicitly represent things in the real world, with relationships between each other that are isomorphic to relationships between real things, and appropriate relationships to inputs and outputs.

Now Searle says "look, this operator still doesn't understand Chinese, but she has the right cognitive processes according to you."

But I have a reply- In this case I'd say that she's effectively been bifurcated into two people, one of which doesn't have semantic access to the meanings of what the other says. When she runs the program of interacting abstract shapes that tell her what to say in Chinese, she is bringing another person into being. This other person is separated from her, because it can't interface with her mental processes in the right way [This "the operator is bifurcated" response is not new- c.f. many such as Haugeland who gives a more elegant and general version of it].

Making the conclusion intuitive

Let me try to make this conclusion more effective through a digression.

It is not by the redness of red that you understand the apple, it is by the relationships between different aspects of your sensory experience. The best analogy here, perhaps, is music. Unless you have perfect pitch, you wouldn't be able to distinguish between c4

and f4 if I played them on a piano for you. You might not even be able to distinguish between c4 and c5. What you can distinguish are the relationships between notes. You will most likely be able to instantly hear the difference between me playing C4 then C#4 and me playing C4 then D4 (the interval C4-C#4 will sound sinister because it is a minor interval. The interval between C4 and D4 will sound harmonious because it is a major interval. You will know that both are rising in pitch. Your understanding comes from the relationships between bits of your experience and other bits of your experience.

I think much of the prejudice against the Chinese room comes from the fact that it receives its input in text:

Consider this judgment by Gary Marcus on claims that LaMDA possesses a kind of sentience:

“Nonsense. Neither LaMDA nor any of its cousins (GPT-3) are remotely intelligent. All they do is match patterns, drawn from massive statistical databases of human language. The patterns might be cool, but language these systems utter doesn’t actually mean anything at all. And it sure as hell doesn’t mean that these systems are sentient. Which doesn’t mean that human beings can’t be taken in. In our book *Rebooting AI*, Ernie Davis and I called this human tendency to be suckered by *The Gullibility Gap* — a pernicious, modern version of pareidolia, the anthropomorphic bias that allows humans to see Mother Theresa in an image of a cinnamon bun. Indeed, someone well-known at Google, Blake LeMoine, originally charged with studying how “safe” the system is, appears to have fallen in love with LaMDA, as if it were a family member or a colleague. (Newsflash: it’s not; it’s a spreadsheet for words.)”

But all we humans do is match patterns in sensory experiences. True, we do so with inductive biases that help us to understand the world by predisposing us to see it in such ways, but LaMDA also contains inductive biases. The prejudice comes, in part, I think, from the fact that it’s patterns in texts, and not, say, pictures or sounds.

Now it's important to remember that there really is nothing qualitatively different between a passage containing text, and an image because both can easily include each other. Consider this sentence. "The image is six hundred pixels by six hundred pixels. At point 1,1 there is red 116. At point 1,2 there is red 103"..." and so on. Such a sentence conveys all the information in the image. Of course, there are quantitative reasons this won't be feasible in many cases, but they are only quantitative.

I don't see any reason in principle that you can't build an excellent model of the world through relationships between text alone. As I wrote a long time ago:

"In hindsight, it makes a certain sense that reams and reams of text alone can be used to build the capabilities needed to answer questions like these. A lot of people remind us that these programs are really just statistical analyses of the co-occurrence of words, however complex and glorified. However, we should not forget that the statistical relationships between words in a language are isomorphic to the relations between things in the world—that isomorphism is why language works. This is to say the patterns in language use mirror the patterns of how things are. Models are transitive—if x models y , and y models z , then x models z . The upshot of these facts are that if you have a really good statistical model of how words relate to each other, that model is also implicitly a model of the world, and so we shouldn't surprised that such a model grants a kind of "understanding" about how the world works."

Now that's an oversimplification in some ways (what about false statements, deliberate or otherwise), but in the main the point holds. Even in false narratives, things normally relate to each other in the same way they relate in the real world, generally you'll only start walking on the ceiling if that's key to the story, for example. The relationships between things in the world are implicit in the relationships between words in text, especially over large corpora. Not only is it possible in principle for a language model to use these, I think it's very possible that, in practice, backpropagation could arrive at

them. In fact, I find it hard to imagine the alternative, especially if you're going to produce language to answer complex questions with answers that are more than superficially plausible.

Note: In this section, I have glossed over the theory-ladenness of perception in this section and treated perception as if it were a series of discrete "sense data" that we relate statistically, but I don't think it would create any problems for my argument to expand it to include a more realistic view of perception. This approach just makes exposition easier.

What about qualia

I think another part of the force of the Chinese room thought experiment comes from qualia. In this world of text associated with text in which the Chinese room lives where is the redness of red? I have two responses here.

The first is that I'm not convinced that being a person requires qualia, I think that if philosophical zombies are possible, they still count as persons, and have at least some claim to ethical consideration.

The second is that qualia are poorly understood. They essentially amount to the non-functional part of experience, the redness of red that would remain even if you swapped red and green in a way that made no difference to behavior, in the famous inverted spectrum argument. Currently, we have no real leads in solving the hard problem. Thus who can say that there couldn't be hypothetical language models that feel the wordiness of certain kinds of words? Maybe verbs are sharp and adjectives are soft. We haven't got a theory of qualia that would rule this out.

I'd urge interested readers to read more about functionalism, probably our best current theory in the philosophy of mind. I think it puts many of these problems in perspective.

Edit: An excellent study recently came to my attention showing that when GPT-2 is taught to play chess by receiving the moves of games (in text form) as input, it knows where the pieces are, that is to say, it contains a model of the board state at any given time. “Chess as a Testbed for Language Model State Tracking” (2021) As the authors of that paper suggest, this is a toy case that gives us evidence these word machines work by world modeling.

The AI Control Problem in a wider intellectual context

Epistemic status: A public intellectual is someone interesting enough that we have decided to let them be obviously wrong. I, unfortunately, am not even a public intellectual.

I've been thinking about the control problem lately. The control problem, also called the AI alignment problem is, per Wikipedia:

[A]spects of how to build AI systems such that they will aid rather than harm their creators. One particular concern is that humanity will have to solve the control problem before a superintelligent AI system is created, as a poorly designed superintelligence might rationally decide to seize control over its environment and refuse to permit its creators to modify it after launch. In addition, some scholars argue that solutions to the control problem, alongside other advances in AI safety engineering, might also find applications in existing non-superintelligent AI.

But can't we just program it to help us rather than to harm us? The problem is that if you give a super-powerful entity a goal- a value function- and it follows it literally- bad things can happen. An analogy- consider a genie. This genie isn't actively malign, but it will do exactly what you tell it to do in the most direct way possible. Wish for a tonne of gold? Well, it appears on top of and/or inside of you because that's the most direct place for it to appear.

Now let me introduce an idea to understand the control problem.

*A thick concept is a concept for which we can check whether any given instance falls under that concept **relatively** easily. However, it is all but impossible for us to articulate rules which, when mechanically applied, will tell us whether a given instance falls under a concept. In other words, it is very difficult or impossible to create an algorithm that captures thick concepts.*

Using our analogy again, we can tell you if the genie has given us our heart's desire (whether something falls under a concept), but we can't give instructions for the genie to follow literally to give us our heart's desire (can't capture it with mechanical rules in a way that won't fuck us over). Ironically I'm not quite sure my definition of thick concept captures exactly what I mean, because later on, we'll look at cases where we can't even agree on whether something falls under a concept, but I think this definition is a good start.

Now let us define a problem, or rather a class of problems. The conceptual richness problems are problems of trying to cope with thick concepts, either by (quixotically) trying to spell them out in all their detail and creating an algorithm, or by finding an alternative to having to spell them out. The control problem is one instance of a conceptual richness problem that specifically arises, at least in part, because there are so many thick concepts in human ideas of the good- flourishing, autonomy, rights, and so on. We can (often) tell you if a computer has respected the good, but not give a computer step-by-step instructions for respecting the good.

I thought that an interesting and bloggable, approach to the Control Problem would be to start a conversation about the variety of disciplines that also face the conceptual richness problem, with the idea of encouraging mutual interchange. Intellectual enterprises that have run up against these sorts of problems include Analytic Philosophy, Classical AI, Law, Statistics in social sciences, and of course AI alignment. Related but separate problems arise in areas as varied as poetry criticism, teaching, and AI-interpretability. Maybe by teasing out the transdisciplinary nature of the problem, we'll encourage cross-pollination, or at least that's my hope.

Analytic Philosophy

Analytic Philosophy has taught us that, save perhaps a tiny handful (and maybe not even that!) all concepts are thick. It has shown this inductively. Hundreds of thousands of person-years have been spent by philosophers trying to find definitions of things-

reasonably compact lists of necessary and sufficient conditions-. No such efforts have succeeded. Granted, philosophers have generally focused on fraught concepts like beauty, truth, goodness, knowledge, causation, etc., but there are no signs that shifting attention to easier concepts would help much. Indeed, consider a paradigmatic example of a concept that is often thought to be easy to analyze:

Bachelor: X is a bachelor if and only if X is an adult, X is male and X is unmarried

Firstly, note that even if this definition succeeds, we've just moved the attention onto three far more fraught concepts, adult, male and unmarried. But secondly, observe that this definition isn't clearly right. Is the pope a bachelor? Is a man in a loving thirty-year relationship with twelve kids who, nonetheless, is not technically married a bachelor? Presumably adult male animals don't count, so we might think it's humans only, but if there were such things as elves, would an unmarried adult male elf be a bachelor? Even this 'simple' term, understood well enough that just about any native speaker could check whether a given use was right, wrong or dubious, cannot be turned into an algorithm.

Analytic philosophers have reacted to the apparent impossibility of finding necessary and sufficient conditions of things in different ways. Some philosophers are still trying to do it. Other philosophers view proposing definitions as a kind of provisional exercise-never fully adequate but useful for a variety of reasons. Others just get on with the many kinds of philosophical work that don't require specifying necessary and sufficient conditions of concepts. Still others are grappling with ideas like the conceptual engineering program in light of these and related difficulties. Work in psychology (e.g. the prototype theory of concepts and related nonclassical approaches) has informed the thinking of philosophers about these issues. Philosophical work (e.g. Wittgenstein's metaphor of family resemblance as a replacement for the idea of necessary and sufficient conditions) has informed many psychologists working on concepts in turn.

Symbolic or classical AI:

I don't know so much about computer science, which is a shame because from what I can tell, problems of conceptual richness abound. It could be argued that they killed (or rather maimed) an entire approach to artificial intelligence. AI wasn't always this machine learning connectionist stuff. Prior to the machine learning revolution, the most promising work in AI was around Symbolic AI. Symbolic AI tried to capture intelligence through explicit representations, operations using rules, etc.

There's a long history of how this approach ran aground- the Dreyfus critique, AI winters, etc. I won't say too much about all this stuff because I don't know it that well, but there's a joke about this sort of approach I like:

"I don't know why self-driving cars keep hitting objects. It should be simple enough to program them not to:

If (going_to_hit_something) Then (Don't)"

Let me translate the humor of this joke. Imagine a robot moving through the world with a camera giving it information about its environment in the form of an array with color data at each point. You are a hapless researcher who has to hard code rules to interpret that array of data into a guess about what the physical space and the objects in it, around the robot look like. Where would you even begin?

Classical AI proved very good at dealing with certain kinds of toy problems, and also with certain kinds of very important problems, like expert systems for disease diagnosis. But most of our ways of relating to the world just proved too thick to capture in lines of code, however extensive. The conceptual richness problem was thus one of the negative triggers for the switch over to machine learning as the dominant paradigm, with a

variety of positive triggers, most especially increasing computational power and data collation.

A good source of further reading on this topic -both on classical AI and on our problem generally- would be the Dreyfus critique. It goes in a similar direction to our argument here, although what we call thick concepts are just one part of it.

Law:

The behaviors we want to forbid and require are complex, varied, situational, come in degrees and are themselves subject to controversy. Spelling out exact rules for judges to apply and civilians to follow might seem impossible, and it is! so in its own encounter with the conceptual richness problem, the legal system has to find alternatives to creating algorithms of law. Unfortunately (or fortunately, depending on who you ask) because law cannot be turned into an algorithm we often face what contemporary legal scholars call legal indeterminacy- a situation in which there is no single right answer to many important legal questions. This has a number of undesirable effects, it undermines the rule of law- the idea that the law should be clear and determined in advance, and hence easy to follow. It blurs the line between judicial and legislative functions- arguably very undesirable in a democracy where the legislature is elected but judges are not (or even if judges are elected, it is difficult for the public to apply democratic scrutiny to their choices).

A big part of the way law approaches the conceptual richness problem is what might be termed constructive ambiguity. Laws are designed as far as is possible to create socially desirable flexibility while avoiding socially undesirable uncertainty. There are many ways to create constructive ambiguity: for example heavy use of concepts such as reasonable. Just add a bunch of steps in the procedure that amounts to saying “refer to best human judgment”. This is why sentiments like “the law is the law” are so silly, the law is full of discretion, and is set up that way deliberately, not that you could avoid it even if you tried.

Consequently, It is often, or perhaps even always, impossible to decide cases without legislating from the bench to some degree. In some cases -maybe even most!- what that legislation should be according to prevailing standards is so uncontroversial that no reasonable judge would disagree. But although humans may agree intersubjectively on the result, that doesn't mean the human judgment is dispensable. Subtlety creeps in.

A great example of this is the (fictional) case of the Speluncean explorers, again via Wikipedia:

"The Case of the Speluncean Explorers" is an article by legal philosopher Lon L. Fuller first published in the Harvard Law Review in 1949. Largely taking the form of a fictional judgment, it presents a legal philosophy puzzle to the reader and five possible solutions in the form of judicial opinions that are attributed to judges sitting on the fictional "Supreme Court of Newgarth" in the year 4300.[a]

The case involves five explorers who are caved in following a landslide. They learn via intermittent radio contact that, without food, they are likely to starve to death before they can be rescued. They decide to engage in cannibalism and select one of their number to be killed and eaten so that the others may survive. They decide who should be killed by throwing a pair of dice. After the four survivors are rescued, they are charged and found guilty of the murder of the fifth explorer. If their appeal to the Supreme Court of Newgarth fails, they face a mandatory death sentence. Although the wording of the statute is clear and unambiguous, there is intense public pressure for the men to avoid facing the death penalty.

I'm going to dispute that last sentence.

The wording of the statute in this hypothetical case is:

"Whoever shall willfully take the life of another shall be punished by death."

It seems pretty simple, right? Willful gives a bit of wiggle room, but on the whole, so long as the deliberateness of the action is not in dispute, the questions of law in murder trials under this statute should be pretty simple. The sentencing phase should be even simpler again! Clearly, on the most natural, direct meaning, the explorers breached the statute. So does this case break down into a Sophie's choice between going with the law and going with morality? No.

Since the publication of the article, a number of legal scholars on both the left and the right have commented on the case. The right often maintains that the law here is clear, and it is not the job of judges to legislate from the bench- even where the law will lead to tragedy as in this case. The left has had a number of lines of reply, but to me the most ingenious is this: none of these conservative commenters have ever thought that the law as written would require you to execute the executioner after he has finished his execution. Yet to the extent the law as written can be said to have a plain and natural meaning, that plain meaning implies that you should execute the executioner, and the executioners' executioner, and so on. Clearly then, no one is taking the law at face value.

So the narrative of what the law plainly says versus external moral considerations breaks down because everyone in the room is interpreting the law in terms of policy goals and ethical values to some degree.⁽¹⁾ Now with that established we are, in words often falsely attributed to Winston Churchill, "Just haggling over the price". The left is willing to stop being literal at a lower bar than the rightist jurists. The leftist jurists continue their argument against the rightists: since there's no Schelling point of literalism to stop at and since we are both being non-literal to some degree - why won't you take a few extra steps to join us? You could stop these more or less blameless men from dying.

My sympathies obviously lie with the leftists here, but it's possible I'm wrong. Maybe in the long run the degree of textual looseness you would need to acquit the Speluncean

explorers is too much. Maybe we should allow enough textual wiggle room to spare the executioner, but not enough to let these men go. But the point is established, I think, that whatever pretenses law might have, it's galaxies away from being algorithmic. The solution everyone has adopted to the problem of conceptual richness in law is just to add a multitude of judgment calls. Some jurists are just more honest about it than others.

Presumably, the drafters of this hypothetical law had something in mind like "I dang think if you take a life your life should get taken". No doubt they could tell you what they wanted in individual cases, but they failed to capture that concept properly (albeit, in this case, it doesn't look like they tried particularly hard). Thus conceptual thickness strikes again.

Statistical queries of a certain sort:

When you are researching a certain type of question, no econometric statistic is ever quite right. An example. Recently I presented some data that showed that, after adjusting for inflation, the average wages of non-supervisory workers and production workers haven't risen since 1964.

Now a bunch of people objected in a variety of ways, one of those ways was:

A) This statistic wasn't quite right because it didn't include non-wage benefits. Others including myself thought it was fairer, on the whole, not to include things like healthcare premiums.

Two objections they did not give, but could well have, were:

B) The statistic wasn't quite right due to Simpson's paradox. That is to say, it is possible that if you divide the workers up by race, each individual race is better off, but due to the changing racial composition of the population, the average isn't improving. I would

object to this objection that it was fairer on the whole just to look at the aggregate if we are to assess the position of the working class qua working class.

C) The statistic wasn't quite right because the percentage of people who are supervisors has increased since 1964, thus those who were "left behind" in non-supervisory roles may represent a less talented pool. My response is that again, we are interested in the plight of the working class, so including non-workers would slant things. A statistician with more resources and time could respond with controls for skills- but this itself would open up numerous debates and questions of judgment.

Step back. What we are trying to capture through statistics is an answer to the question "has the American working class had an unusually bad period, economically speaking, in the last 55+ years". But that question contains a series of thick concepts- e.g "unusually bad period", "economically speaking" "American working class". Because these concepts are so thick, it's all but impossible to design a single statistical query with exact parameters that captures perfectly the intuition behind the question.

Partly the answer in statistics is sensitivity testing- many different queries with different combinations of parameters to see if they all lineup. Partly the answer is a judgment call- I think that the statistics I gave were, broadly speaking, very fair. Once again though we've hit the conceptual richness problem, our concepts are too broad and subtle for what we can capture in a formally defined query.

Disciplines that have grappled with related but separate problems include:

Poetry criticism:

Poetry critics have long lamented or rejoiced that it is impossible to capture the full meaning of a poem through criticism. Harold Bloom once wrote that the meaning of a poem could only be another poem. Now a lot of STEM types would probably dismiss this as lunar-eyed romanticism. I propose though that we take it seriously. A poem (or

any artwork really, but it's especially clear with poems) creates a kind of mental experience that is too rich to spell out. Obviously part of the problem is that to experience a poem is to feel something, and you can't usually explain someone into feeling something. However, I think it is entirely plausible that another part of the problem is this: experiencing a poem means experiencing something too complicated to be explained systematically. I see this as having kinship with the conceptual richness problem, although it's not quite the same thing.

Teaching:

In many sorts of teaching, e.g. teaching about complex concepts like "alive", we can often tell the student whether x falls under that concept, but we can't give the student a rule. Often the compromise in teaching seems to be giving the student an approximate rule and advising them that there are exceptions. Eventually, through a poorly understood process- a meeting of inductive biases and experience- the student gradually gets it. In the words of Wittgenstein: "Light dawns gradually over the whole". Once again, this very general and common situation seems analogous to the conceptual richness problem.

Machine Learning Interpretability:

At the moment in artificial intelligence there is a great deal of attention being spent on the problem of interpretability. Machine learning programs trained on millions or even billions of examples can use this knowledge to very good effect- sometimes better than humans- and it would be nice if we could use these programs in place of humans sometimes. The problem is that for all sorts of reasons, we can't do this unless computers can explain their choices. In some sense this is an instance of our problem- what we would really like is for the computer to take its complex, statistically layered concepts and applied processes and translate them into reasoning a human can understand. This sounds quite similar to the problem of translating human concepts into algorithms. Obviously, this isn't quite the same problem as the problem of distilling

concepts, because humans can give rationales for their decisions but can't distill concepts, but I suspect the analogy is important.

Incidentally, I wonder if the machine learning interpretability problem suggests a skeptical possibility about human communication. Maybe we make our decisions on the basis of vastly complex processes that bear very little resemblance to the explanations we give for our decisions. Maybe all or nearly all explanations are just post-hoc rationalisations.

Summary:

1. The control problem is an instance of, or is at least very closely related to a very general problem. Simply put, that general problem is that we can use our concepts, but we can't understand them in a systematic, formal way.
2. To the best of my knowledge, this problem has never been given a domain-general name. I call it the problem of conceptual richness.
3. The problem is likely insoluble in the way we would most like to solve it: humans writing out a procedure, which could be mindlessly or near mindlessly applied.
4. But there are alternative approaches to coming to grips with the problem. Exploring how other disciplines have approached this may be an interesting direction in the study of the control problem.

Footnote

(1): Here's one unconvincing attempt to get out the problem: "we need to look at what the drafters intended, and not just the literal meaning. Clearly, the drafters did not intend that the executioner be put to death, but clearly, they did intend that people like the Speluncean Explorers should be put to death".

Certainly, I agree that the drafters clearly didn't intend executioners be put to death, but in what sense is it clear that they did intend explorers to be put to death? It seems entirely possible that they'd be horrified by that reading. In truth the drafters probably weren't thinking about cases like the explorers, so they didn't intend anything either way on that sort of case. Drafter intention will fare no better than plain meaning, or only a little better. Plus the epistemic difficulties in getting to it are much greater, but that takes us beyond the scope of this essay.

Fiction, Criticism & Similar

Just read the Damn Book of All Hours- my (losing) entry for Scott Alexander's book review contest

Note: This review is of both “Vellum” and “Ink”, books one and two of the Book of All Hours respectively. However, I only had time to reread Vellum in preparation for the review, so the focus will be on Vellum. The title of the duology is “The Book of All Hours” and I feel that, spiritually, Vellum & Ink are a single novel, and were probably intended as such but split for commercial reasons. Thus I sometimes call it a book in the singular.

Spoiler warning: This review contains spoilers. I would be very surprised if, given the nature of the book, spoilers spoiled anyone’s fun. At the very least I am confident that these particular spoilers won’t.

“The annuna, the judges of the underworld, surrounded her- They passed judgment against her. Then Ereshkigal fastened on Inanna the eye of death- She spoke against her the word of wrath She uttered against her the cry of guilt- She struck her. Inanna was turned into a corpse- A piece of rotting meat And was hung from a hook on the wall.”

-The Descent of Inanna

There are some house words at the top of Scott’s blog that say “in a mad world, all blogging is psychiatry blogging”. True, and in a narcissistic world, all reviewing is, at its heart, self-review. So we come to Vellum & Ink by Hal Duncan (2005, 2007) and my body and blood. We come to what it says about me that I consider this obscure fantasy duology the greatest fiction I have ever read, bar nothing. Perhaps such an idiosyncratic judgment is more of a haruspex of my innards than the book, but still, let me show it to you. I’ll be direct, I love it, I think it’s glorious and I’m going to try to convince you to read it. I’m probably going to get overexcited, definitely cringe. I apologize- I can only plead honesty. I hope I succeed in getting you to pick it because you guys would love it-

mythology, erudition, word games, psychology- it's like an inventory of all that this Substack enjoys.

How to give a flavor of it?

“A story saying that all but one solitary page were blank, and on that page, there was only a single simple sentence, an equation which captured the very essence of existence. This, he said, was why all those who'd ever looked upon the book had gone insane, unable to comprehend, unable to accept, the meaning of life laid out in a few words of mathematical purity. After what happened to Thomas, I remember thinking that I knew what that sentence was. Two words. People die.”

But whatever is written in the book “it is not in heaven”. There's always hope. No judgment is so ineluctable that it cannot be run from, reinterpreted, resisted, revised, rejoined, reimagined, wrestled against or, worst-case scenario maliciously complied with. Thus, at least as long as there are still people, people have a chance. This is a story of hope militant and armed. The story of the clash between these powers- the certainty of death, the fire of hope- on the battlefield of eternity.

I- When I found the doors of perception out through the gates of heaven and hell

A quarter way through this, the journey of our mortal life, I received book vouchers valued at 25 dollars for a science fiction and fantasy bookstore from a friend. It was on the occasion of a birthday, I think my 21st. So I went into the central business district, planning to browse and then impulse buy. I can't recall exactly, but knowing me I was probably looking for something science fiction or fantasy- with pretensions of being a wee bit literary- think *The Use of Weapons* or *The Raven Tower*.

I found *Vellum*, Book one of the *Book of All Hours*.

On the back, in the sonorous language of blurbs:

“The Book of All Hours: I

In the Vellum, – the vast realm of eternity on which our world is just a scratch – the unkin are gathering for war.

In the Vellum, – a falling angel and a renegade devil are about to come to blows.

In the Vellum, – blood magic made in hell is about come face to face with nanotechnology forged in heaven. Past, present and future will collide with other worlds and ancient myths.

And the Vellum will burn

An extraordinary, incendiary masterpiece from a rare new talent, this is multi-stranded, multi-charactered imaginative fiction that blows traditional literary concepts apart”

The blurb is a bit cringe but what blurb isn't cringe? All blurbs sound like they're meant to be read by that guy who does all the movie trailer voiceovers. I started flicking through it and decided to buy it. I think I recall that I was so confident in my choice that I bought the second volume at the same time.

I got home and I started reading it. I became confused. Surprisingly, it did blow traditional literary concepts apart, at least for a certain value of traditional. It reads as if James Joyce got into Kabbalah or like Scott had written *Unsong* while high on LSD. I became very lost. So I went and grabbed a notepad and started from the beginning, drawing little diagrams and keeping little glossaries as I went. I started to grasp it, indeed, it seared me.

II- The lost days of summer and the lost deus of Sumer

Trying to summarize this book is a bit like trying to capture the essence of a complex three-dimensional object like a human body with a single two-dimensional slice. Nonetheless, the simplest way to describe *The Book of All Hours* is that it tells a story not about characters in the normal sense, but about hyper-characters- archetypes. It does this by weaving a series of stories, with the same archetypes in them, together into sequences corresponding to a narrative of narratives.

The difficulty in reading it is peculiar, probably not quite what you are expecting. You know how Cthulhu is so maddening to look at because he is an N-dimensional being compressed, with loss, into a 3-dimensional space? It's like that kind of difficulty, much more startling than it is laborious.. Or to look at it another way, It has been said that Kant is difficult to read not because he is vague, but precisely because he is so exact. Something like that is happening here. Given the architectonic character of the book, I was not surprised to learn the author is a former programmer. I can only imagine what the storyboard must have looked like.

The “center of gravity” of the stories is that the heroes- these archetypes moving through the space of all possible worlds-are Unkin. Once-humans who have stepped out of linear time, into a space of narrative time, they became demi-gods. Over the millennia they have been arranging themselves into two sides: The Covenant- an attempt to govern the Unkin, to bring them to heel and cease their pretensions to Godhood- they have championed Abrahamic monotheism. The Sovereigns stand against them- individual self-styled deities and their little fiefdoms. Angels and demons duel with swords of fire and words of thunder in the air. But a third side is coming, a Greek chorus that has chosen to enter the fight, a swarm of dead souls, moved by the richness of the world out from death and eternity, furious at the pretensions of both sides:

“NO MORE VILLAINS, NO MORE VICTIMS, NO MORE HEROES”

[...]

A commendable concept says Moloch bitterly. And who might you be?

And we tell him exactly what we are”

A searing flash of light is seen across the skyline of Damascus...”

Every war has its deserters. Our lead characters are deserters from the war in heaven, trying to avoid being pressganged into either side. One of the characters, in various folds of reality, goes by the names Puck, Thomas, Dionysus, Tammuz, Dummuzi, Adonis, Pan, Thomson’s Gazelle (the species) Matthew Shepard (yes, that Matthew Shepard) and many others. He is killed, always by at least two killers, one of whom who admixtures disgust and longing- Jack. At the timeless, omnipresent moment of his death, a cry is heard across beingscape.

“The Great God Pan is Dead”

As was reputed to have been heard in our world, during the reign of Tiberius.

The characters, and the hyperspace of reality, are rent by grief. The multiverse starts circling the drain. Why does the death of Puck cause so much harm? No one is quite sure, and the characters themselves remark on the strangeness. However, within the dream logic of the book, I think a better question might be why doesn’t the death of anyone destroy the multiverse, for as it is written in the Talmud & Quran: Whoever destroys one life is as if he destroyed a whole world, and whoever preserves a life is as if he preserved the whole world.

Our heroes wage a desperate struggle -military, political, artistic, sexual, religious, social-against all the laws of narrative and reality, a war against the war in heaven, to restore him to existence.

III- The Ivory Tower weighs in

There might be only three academic works on this book. Two I have been able to find: *Rewriting Myth and Genre Boundaries: Narrative Modalities in The Book of All Hours* by Hal Duncan by Popov (2020) and the earlier piece *Dead gods and rebel angels: Religion and power in Philip Pullman's His Dark Materials and Hal Duncan's The Book of All Hours* Macaskill (2009). I also found references to another work *Euripides Bound: Hal Duncan's use of Greek Tragedy* (2009) but couldn't find a copy. The abstract of Popov's article might give a flavor of the absurd but never quite ridiculous enterprise of the book:

This article explores the relations between fictional time, genre, myth, and narrative modalities in Hal Duncan's novels *Vellum* (2005) and *Ink* (2007) – known collectively as *The Book of All Hours*.¹ [...] The metanarrative mechanisms of the novels are then explained using this apparatus, which demonstrates the usefulness of SF in challenging fundamental assumptions about the grammar of thought. Finally, the same theoretical and methodological approaches are considered as tools for expanding narrative engagement with the world beyond strictly human domains, connecting the overarching argument to works from recent theoretical developments such as Object-Oriented Ontology and agential realism.

A pretty ambitious article for a book everyone has forgotten! I haven't read the article cover to cover, but he's right about one thing, it's impossible to read the book and not think about possible worlds and long afternoons debating whether Twin Earth water is water.

IV- Of power and its false opposition

I once wrote to Hal Duncan and suggested that, what the book is about more than anything, is an attempt to escape both power and its false opposition. Hal Duncan agreed with that. And who amongst us has not looked at the webs we've wrapped ourselves in, look at the grandeur and potential of humanity, and not found ourselves saying "Christ, but do you have any more options for me?". It might seem like only a child's lament, a refusal to grapple with the Serious and Hard Choices. But consider all the dilemmas- technological, social, ethical, of the past. Think how often the right answer later turned out to be some "none of the above" a third option that hadn't yet been imagined.

V- Yearning

Let us define yearning in contradistinction to desire. Desire is for something concrete. Yearning is for something you can't capture in words or pictures. Desire is GPS coordinates, yearning is "Second star on the right and straight till morning".

Why did this book sear itself onto me? The yearning. For so long I've yearned to be an abstraction, rising above a particular life, to become an idea-force operating on history. I've dreamed of writing the perfect book that summarizes exactly what I think and then just fading- leaving the book behind. Becoming an archetype, disincarnate agency. This book speaks to that.

However, I've also dreamed of becoming concreta, filling up each moment to overflowing. I've dreamed of being like Thomas sitting under the apple tree on a late summer afternoon, dappled in light, even while in his past and future both he is pursued and dying. Still, this moment is perfect for Thomas, so all moments are.

-You'll always be getting captured, she says

-And I'll always be escaping them

-They'll kill you over and over again

-And all the time I'll still be here, he says,, under a tree of golden apples and green leaves

The Book of All Hours speaks to both longings and so many more. I've yearned for things that don't have a name. I've yearned for seemingly logically contradictory states of affairs. I have longed to be ordinary and longed to be extraordinary, longed to help the world and longed to escape from it. Haven't we all? There is nothing special about me in this regard- everyone yearns. But this book gets yearning. Big deal- don't a lot of books get yearning? Sure, but how many so successfully echo the infinite dimensions of yearning in their form and content?

It is a book that resonates in the hollow places of those who have loved beyond reason and shook their fists at death. In its careful balance, it neglects neither the sweetness of life, nor the longing for transcendence.

And as his cathartic inferno lights a slant of angular face, his obscenities, his profanities, his blasphemies turn into sobs and laughter, invocations of his lost love that break my heart as I sit here in the car, the engine idling, my hands shaking so much I know I cannot write those actual words of rogue desire without dissolving.

I think the book has turned me into a romantic in a lot of ways. For example, I am a utilitarian, in that I want to maximize welfare, but I could never agree with those who think welfare is simply the balance of pleasure minus pain. True welfare is, I think, a kind of wholeness of being and mastery of capacity. Pleasure or desire satisfaction might approximate it, but only approximate it. Intellectually my reason for thinking this is that, if we were pleasure maximizing utilitarians, we would think that a universe stuffed full of people reliving their greatest experience over and over again would be "good". In terms of a philosophical autobiography though, I think it was probably this book that awakened me to this dimension of life.

So often, we describe all our values, especially, but not only our political values, in a stationary way, in terms of a possible utopia. Maddening! Our values are not places, they are directions. They are yearnings. In our limited minds these yearnings cannot be articulated ahead of time, our experience of trying to get someplace better, in life, in politics, in art, is much more like following a star than it is like following a map. Maybe this is only true because of human limitations but it is true.

VI- Verdant prose for the lost deus

The prose is intricate and luxuriant yet made according to his careful design. Ornamentation and function blend together. Each sentence picks up speed as it goes along, accelerating. Each sentence is a self-contained world, even when short. They are each laden with significance, but, at least in context, rarely pretentious. In parts, it is almost a verse novel. All this is all the more astonishing given it was Hal Duncan's first.

Let me pick out a selection, almost at random, by opening Vellum up to different pages.

“There is a city, canopy, he says, last of a distant land and near the fountains of the sun, inhabited by a dark race”.

“The Jornada del Muerto, the Journey of the Dead Man, runs from Kern's Gate, El Paso, north through a dry plain of natron, uranium, salt, sand and dust, up to Santa Fe, up to Los Alamos and Trinity where they shattered atoms, those destroyers of worlds”.

“The birdmen who sing the morning world into existence with their cant”.
(Birdmen=angels, cant=language of magic).

““Everything is real” said Jack “Everything is true; nothing is permitted”.”

VII- Grit among the angels and light in Stygia

There's a sensuality in the book, empiricism that prevents it from becoming a metaphysical treatise. Sometimes, repulsive, sometimes alluring, sometimes both:

“Rotting metal, petrol fumes, boiled blood, sulphur or ammonium. I swear it smelled like all of them and none of them. If that smell is natural than nature is no mother”

But the sensuality isn't just in the words, it's in the rhythms, the flow of speech which forces you to process the book audibly, and not just in terms of its semantics. Whole chunks are in rhyme, sometimes completely out of the blue. In the ideal case, this book would be read aloud, cover to cover, but a more select approach of reading the passages - the songs, the poems, the interludes by the Greek chorus- that vibrate in their sound will suffice.

IX- A world built of signification

The layers of apophenia and parallel go very deep, they are structure rather than ornamentation, and cannot be resolved into a single coherent symbol system. For example, which of the main characters represents Jesus? Is it Jack Carter (initials JC)- a redeemer born to overthrow tyrants? Is it Seamus/Prometheus- stretched out on a rock for the sins of humanity? Is it Thomas/Puck? The God who dies for our sins yet eternally lives again? More speculatively, is it the cold-blooded Joey?

Joey=Yoey=Yoo-eee=Yod-He-Vav-He made flesh? A more advanced reader than me could probably make a case for any of the book's seven characters being a stand-in for the messiah. Moreover, because the characters are archetypes rather than singular characters they constantly shift relations- is Anna Thomas's sister? his Mother? his Lover?

The books reward rereading. At first you uncover the layers and hints left by the author. Every time you read it you will discover new hints, but gradually you will begin to build your own structure, stake out your own claims in the space of possibility.

I would very much not recommend this book to someone in imminent danger of psychosis. As someone who went pretty far down that path when I was a small child, it walks the road- makes a “steelman” of psychosis- all too well:

“You hear voices Jack?”

“Don’t we all? Voices of souls, of ancestors, family and friends, enemies and demons, ghosts inside the head, the ghosts in the machine. You telling me you don’t hear your own little internal narrative when you’re thinking to yourself? You’ve never had an argument with a friend that didn’t carry on in your head afterwards? You’ve never lain in bed and thought to yourself in someone’s voice, to get a different perspective, someone else’s attitude? We all hear voices, doctor. Most people just keep them turned down real low.”

“And these voices tell you to...”

“Listen. It’s like being asleep beside a river, a river of voices, babbling, buried in the rustle of leaves. Narcissus sleeps and dreams us all.”

X- En Passant

I should mention, at least in passing that the book has a rich and understated British humor that only becomes apparent once your eyes have adjusted to the other lights that dazzle you. Some of them are pretty obvious- e.g. Moloch talking about his “Philistine Liberation Organisation” but there are many more, often revealed upon rereading.

Also in passing, it's a great book to beat people who sneer about genre fiction over the head with. It's more erudite and urbane than just about all of the "sad-middle-class-people-having-affairs" set.

Now comparisons- what's it like? It's a little bit like A) *The Wasteland* by T.S. Eliot B) *Ulysses* by James Joyce C) *Unsong* by Scott Alexander D) Any book from a genre with "punk" at the end- steampunk, sailpunk, solarpunk E) The works of Kirkegaard. F) Reading *The Golden Bough* by Frazer while high on LSD.

XI- Sodomy and violence

In understanding the book's failure to take off, most of the blame must surely lie in its difficulty. However the choice of the author to make the story, in some sense, about a gay romance doubtless didn't help. Just how many people are there who want to read a gay romance, written in the style of James Joyce, against the backdrop of the space of all possible worlds? Perhaps if it had been written today the thriving slash community would have carried it, I'm not sure.

But in other ways its publication today would be a little more fraught. Jack is sometimes Thomas's killer, sometimes Thomas's lover, and sometimes both. A deliberately disturbing choice. It's a dynamic that would be difficult to explore with a heterosexual couple without provoking outrage. Perhaps it would be difficult to explore with any couple in a post-Metoo era- see the controversy over Lana Del Rey's "Ultraviolence". But it does make for an entrancing meditation on the erotics of violence.

These days it's very unfashionable to suggest that homophobia might reflect latent homosexuality, and it's very unfashionable to eroticise homophobia. Maybe because of its time, *The Book of All Hours* does both, and whether it is right on the psychological facts, is a richer and certainly more provocative novel for it.

“I started to remember it all at the funeral. I was thinking about how we first met. It was in this social studies class. We just clicked, like we’d known each other all our lives. At the funeral, I started to remember that we had. All our lives. We were like children playing in the illusion fields, he continues. Let’s pretend. One day we’d be soldiers in the First World War- Captain ‘Mad’ Jack Carter, Private Thomas Messenger. Another day I’d be a seraphim sent to hunt him down across this weird graey version of America, only to wake up in a town called Endhaven, amongst black-suited refugees from a nanotech apocalypse. We’ve been shepherds in Arcadia and rent-boys on the streets of Sodom, Doc. I’ve crossed deserts wider than the world because he dared me. I’ve led armies to destruction because he was in danger. I remembered it all- Christ, it was like being born again- as I was standing there at his funeral, listening to the Minister spout his bullshit. We’ve lived a million lives and ended up together, whatever fold of the vellum we were in.”

“And in these other... folds, Puck didn’t die?”

“Oh, no, says Jack. He always dies. You should know Reynard. You should remember too.”

XII- Kill your gays?

On the topic of homosexuality, the TV tropes wiki has a useful comment on the books, which I have often found myself contemplating. The “Kill Your Gays” trope is a postulated tendency -and it accords with my observations although I haven’t seen a quantitative study- for gay characters to be particularly likely to die in fiction. In the bad old days, this was because gay characters were villains and needed to get their comeuppance. These days it’s often the opposite. The gay characters are portrayed as dying because they are too good for this sinful world. Nevertheless, they die.

TV Tropes suggests that the Book of All Hours is an attack on that trope. An attack on the idea that it is normal to expect gay characters to die gracefully and early in a nice

little tear-jerker. But- how can that be? Isn't the book an instantiation or even apotheosis of the trope? Isn't Thomas always the wreathed sacrifice across all the worlds? In a way, but remember, this is a narrative about narratives. Thomas dies everywhere because that's what happens in existing narratives. But the characters are in rebellion against the archetype, trying to find or create a new manifold of possibility where this isn't true.

We could even- and I hope this doesn't paint Hal as too self-indulgent- suggest that the activity of these characters in trying to make a new Eden where Puck can live is a metaphor for the work of the author. The author, or at least Hal, fights against the "tradition of all dead generations weigh[ing] like a nightmare on the brains of the living". Hal invites us to imagine/demand other moral economies in literature where gay people don't just elegantly die, thoughtfully providing a tasteful frisson of sad-spice to make the work Literary with a capital L.

XIII- An invitation to sit in the dappled light under the apple tree on this late summer day

It's sad this book is all but forgotten. Maybe it would have done better today, with the internet and Wikipedia more ubiquitous, allowing the research-as-you-go detective puzzle box approach that this book benefits from. This is a book that really needs to be read by a community. It should be solved together like an alternate reality game on a Reddit forum- except less linear than an ARG, more like a continent to be explored in a series of joint expeditions.

At least until machine learning was invented, writing a book was the closest we could come to making a person artificially. Reading a book is always the interpenetration of two worlds, you and the book. I wish I could give you not just this book to read, but the interaction effect it had on me. But it's for the best I can't because you'll build your own collided world. I hope it will be vast- full of ancient forests and singing deserts- but

regardless, it will spin into fractal slices you can't even imagine now. Or maybe you'll hate it. No way to know without trying.

Pick it up and come sit with me under the apple tree.

The concept of cringe is cringe

Imagine a dude who carefully assesses every aspect of his rhetoric to make sure it doesn't sound close to something a naive, overly emotional or culturally unsophisticated person might say- in other words, he's hyperfocused on not appearing adolescent or jejune. A little bit of this might be healthy but suppose this man internalizes the game so heavily that it begins to shape not just his rhetoric, but his beliefs as well. One of his primary ways of arguing ends up being insinuations that the person he's arguing with sounds like this or that type of annoying person from high school.

I would suggest that, ironically, such a person would resemble very closely an adolescent. There's nothing so impossibly adolescent as being obsessed with not being an adolescent. (Also he's a bit of a Patrick Bateman, but that's another matter.)

Consider, for example, atheism. Right now on Twitter there's a micro debate going on about whether religion or atheism is more cringe. A typical example of the dialectic:

“religious people are *usually* cringier than atheists”

“Hard disagree. What's cringier than guy scoffing about Flying Spaghetti Monster or just-asking-questions about Sky Daddy?”

“I gave my child MMS (bleach) so that I could cure their autism through the purity of our lord” Far more cringe. Dramatically more cringe. We dont have to bring up YEC even.”

“atheists are typically redditors so this is patently false”

“A guy who thinks “sky daddy” is a funny joke vs The “monster energy drinks are satanic” lady Epic cringe off”

Those who hold that atheism is more cringe produce classic prooftexts like this cospypasta from the atheism subreddit back in 2013:

"Just to be clear, I'm not a professional 'quote maker'. I'm just an atheist teenager who greatly values his intelligence and scientific fact over any silly fiction book written 3,500 years ago. This being said, I am open to any and all criticism.

'In this moment, I am euphoric. Not because of any phony god's blessing. But because, I am enlightened by my intelligence.'"

So a kid posted something on the atheism subreddit in 2013 and this is meant to reflect on whether atheism or religion is more uncool, with the unstated but clearly implicit idea that this should affect your religious beliefs, or at least your public profession on religious belief or lack thereof.

This is not at all atypical. Similar fights break about the cringe factor of ideologies, genders, sexualities, and hobbies. Twitter is obsessed with cringe, and although Twitter is largely politically and economically powerless, it does, to a degree, lead the charge on culture.

But I have to object to all this worrying about cringe. You know who spends a lot of time arguing about what is cool and what isn't? You know who is deeply concerned about looking uncool? Deeply uncool, cringe people. We've reached the point where the only thing to do is to declare the concept of cringe, cringe.

Now there's often a rational kernel to this stuff. The rational kernel to the idea that atheism is cringe is that making a big deal out of not believing in God, absent any larger social program or critique of social conditions, can get pretty self-indulgent. What's the point of frantically telling everyone The Lord doesn't exist if you're largely satisfied with

the social structures generations of believers created? If it's just a point of metaphysics to you, why are you pushing it so hard?

The rational kernel to the idea that religion is cringe is that sometimes (not always) religion is used to avoid facing our lived world "as it is" and grappling with the needs and personhoods of others.

But a thoughtful atheist and a thoughtful religious person are both infinitely less cringe than a dude- religious or not- who cares deeply about whether it is atheism or religion makes a person look more naive, unsubtle, adolescent and cringe. Think things through like an adult, and talk to other people like an adult, enough of this highschool clique ranking nonsense.

TLDR: Worrying a great deal about being cringe is self-defeating because it's extremely cringe, but even if it weren't, don't you owe it to yourself to think about serious questions in a serious way?

The Culture Novels and the deaestheticisation of politics

I. All is (not) as it seems!

You probably know this hoary cliché in science fiction. The hardy explorers find a society that appears to be a utopia- but not all is as it seems!

In many ways, The Culture novels by Iain Banks are an inversion of that. Characters either live in, or encounter, The Culture- a society that appears to be paradise and really is. However the protagonists rebel against it- seeking a world where maybe things are a bit tougher, but damn it, there is glory! They convince themselves that the Culture is a fake utopia. Tragedy results.

Characters embodying this trope to a greater or lesser degree are especially prominent in the early books and include:

Horza from *Consider Phlebas*,

Gurgeh from *The Player of Games* (long fantasises about joining contact to escape ennui although he has to be pushed in the end)

And Zakalwe in *The Use of Weapons*.

It runs throughout the culture series, but I think is especially pronounced in the first few books. A case could be made for many others as well, for example Ziller in *Look to Windward*, although he is content mostly to complain.

Some of them oppose the culture altogether, like Horza, others like Gurgeh want to go somewhere else for a little while, imagining great adventure. Some of them die, others escape relatively lightly with a large to moderate to degree of trauma. All suffer because

they want existence to have more friction than the Culture offers. Implicitly or otherwise, they romanticise suffering only to realise in the throes of suffering that there is little romantic about it.

“No, Andrea: Unhappy is the land that needs a hero”

-Bertol Brecht, *Galileo*

II. Utopia as a problem for the writer

It is difficult to tell a story in a society in which there are no objective causes of suffering. No risk of violence, no ill health and no risk of poverty, not even a risk of unhappiness unless you are unwilling to use the appropriate chemical correctives. Even boredom is banished for all but the most jaded tastes. There are a few possible stories that remain: comic sitcoms or romantic comedies, tales of unrequited love or a memoir of the grief of those left behind when someone voluntarily dies. These are real possibilities to be sure, but far fewer than the stories one can tell in a world like ours, still bound up in the realm of necessity. Banks for his part doesn't even try to tell a “purely Culture” story. Every Culture novel is about the culture encountering other societies that are not so utopian.

I think these difficulties are the reason why so many authors instinctively rebel against utopias, and why there are so many fake utopias in fiction which are unmasked as hidden dystopias. Authors instinctively don't like utopias, because they make bad stories. Thus fiction writers, consciously or otherwise, judge the sociopolitical structure of societies on aesthetic grounds, equating poor soil for narratives with poor soil for human flourishing- in the words of Walter Benjamin, they aestheticise politics.

III. Aestheticisation and fascism

It's a shame that authors instinctively aestheticise politics, because this reflex, as Benjamin argues, is one of the wellsprings of fascism.

This was true in the time of the original Nazis and its true of the cheap knockoffs we have now. If you've ever seen Nazi memes they're an ideological mess, but one of the themes that comes through is an obsession with beauty and good narrative as a political goal. From talking about "the ancestors" (always romanticised beyond recognition), to the content of the fourteen words, to bromides about honor that clearly come from the pages of boys own adventure books (and of wars imagined), to complaints about "decadence" which basically boil down to "I don't like looking at it".

In this regard I do not mean to suggest that a bunch of science-fiction and fantasy authors are closet fascists, I'm sure they're good centre-leftists and all that. I'm not judging them either- it really is very difficult to write a good story set in utopia. Nonetheless, we must recognise some basic truths. Suffering, want and involuntary death are bad, and if we do not destroy ourselves, or permanently prevent our own technological progress, we will one day abolish them. Our sharpest want and most urgent action should be to speed this day. Romanticising suffering is cosmic Stockholm syndrome, boot-licking for the brute forces of the universe. Portraying utopias as really secretly evil is a lazy and overdone trope, but, moreover, it values a certain kind of narrative satisfaction above our soaring possibilities.

In the words of Belinda Carlisle "Heaven is a place on earth".

Hypotheticals

I love hypotheticals. I think they're an integral part of the life of the mind- from learning to reason practically, to learning to separate concepts. Here are some hypotheticals that have buzzed through my mind over the years.

Superpowers

- You can read the minds of people around you- in fact you find it difficult to stop. Only with uncomfortable effort you can prevent yourself from reading minds. Do you have a moral obligation to constantly make that effort? Assume that simply disclosing your powers to everyone you meet is not an option, because you're worried you'll be experimented on by the government if they find out.
- You always know the exact words to say to get what you want from someone. If there are any words that would persuade a person to do x- where x is what you want them to do, you instantly know those words and how to say them. What are your obligations in using this power?
- God assures you that your big fears will never come to pass. You will die long in the future peacefully, with nothing horrific happening to you in the interim. How does this change how you live your life?
- Similar to the above. God tells you that you're going to have a great life in every respect but one- you're never going to have a long term romantic partner, no matter how hard you try. How does this make you feel?
- Once only, you can send a 50 word manifesto to everyone on planet earth. Everyone will immediately hear it read in their mind, when doing so would be safe (e.g. not while they're doing something dangerous). They will understand it, even if they don't speak your language. What do you say? If you refuse to send one, why?

Personal choices

- *Applicable only to heterosexuals and homosexuals.* You can never have sex or romantic intimacy with your preferred gender ever again. Do you pursue your non-preferred gender, or accept a life of abstinence?
- Would you rather have almost no self-awareness or too much self-awareness?
- You are given two options by the fates. 1. You can spend the rest of your life with an amazing person, in a relationship filled with great romantic intimacy. Also you'll be having sex with very attractive people. However, the people you're having sex with won't be your romantic partner- that relationship will be sexless. If you want to have kids with your romantic partner it will have to be via artificial insemination 2. You can be in a relationship, with both sex and intimacy, with a somewhat less amazing but still lovely person. Which do you pick? Do you sever sex and intimacy, or keep them together at a loss to the quality of both?
- Think of three very different things you'd be willing to sacrifice your life for.
- Do you believe that relatively small changes in your life or upbringing could have made you do terrible things?
- What's the smallest thing you'd give up your right to vote for? If you refuse to vote, what's the smallest thing that could make you vote? If you can't vote (e.g. resident alien, felon etc.) what's the most you'd give to be able to vote?
- Someone offers you the following deal. You'll die, but a book containing your basic view of the world, most important thoughts etc. will be published, and read by half a billion people. Would you take it?
- From a purely selfish perspective, not considering any larger impacts on the world, which Nobel prize would you most like to win? Options: Physics, Chemistry, Medicine & Physiology, Literature, Peace and Economics.

Ethical odds and ends

- You can save one kind of charismatic megafauna (maybe Pandas, maybe blue whales) or thirty invertebrate species that, while they're not much to look at, biologists tell you are absolutely fascinating from a scientific point of view. Assume no flow on effects like ecosystem destruction or cascading extinctions. Assume also that there will be no lost scientific discoveries that will save or improve human lives. It's purely about the level of scientific interest and the value of species richness, versus human interest in megafauna. Which do you pick?
- In the future it's possible to assemble sex-androids that look and act almost exactly like the real thing. Even assuming you keep it absolutely secret, would it nonetheless still be unethical to construct a sex-android in the likeness of someone without seeking their permission?
- You can live a quiet, mediocre life with almost no impact on anyone else, good or bad, or you can live a wonderful life in Nozick's experience machine. Which do you pick? Be honest.
- You're dead. Gazing from the afterlife you see that your sister/brother is marrying your widow/widower. How do you feel about this?
- You have to ban one and only one forever- either porn or trashy romance/erotica.

Religion

- If you are not religious, is there any conceivable event which would make you a believer? If you are religious, is there any conceivable event that would make a non-believer?
- You find out the following things to be certainly true: 1. There is an all powerful, all loving God (assume some satisfactory solution to the problem of evil) 2. History is the unfolding of a plan by this benevolent God. Think

through, in detail how this makes you feel. If you already believe the above to be true, what about the opposite? You find out there is definitely no all loving all powerful God who acts providentially in history. How do you feel?

Truth

- A working, relatively cheap, lie detector is developed, how does it change society?
- How does this lie detector change humanity's view of humanity?
- You develop the following power- so long as you are being honest, people cannot doubt that you are being honest, although they might think you are mistaken. Everyone knows you are being honest, and everyone knows that everyone knows you are being honest. What do you do with this power?
- You can press a button. If you press this button, the moral character of every single person on earth (defined as altruism divided by selfishness) will be instantly visible to anyone who looks at them. Do you press it?
- You can press a button. If you press it, everyone who is in the bottom ten percent of the population for (altruism/selfishness) is revealed as such. Everyone will know them just by looking at them. Do you press it?

Punishment

- Everyone who ever lived has been digitally reconstructed using technology. You're somehow in charge of the project and have great discretionary power in relation to it. A group of people, calling themselves the wronged, demand that people who did wrong in life should be punished in some way. The injustice of seeing those who hurt them living comfortable afterlives stings them. They differ among themselves on who should be punished. Some want only the worst of the worst to be punished, others want all murderers, torturers, rapists etc. to be punished, and some want to go even further. Adding to the confusion

is the vastness of cultural space represented by all 120 billion humans who ever lived. What do you do, and how do you justify it? Having the totality of people make a democratic decision about it among themselves is an option, but if you pick this you need to justify it, and think through the procedure by which the decision would be made.

- Let's say that there were a condition, call it demonic possession, that could make people do awful things. Let's say, further, that science has demonstrated that this condition exists, is common, and is in no way the fault of the sufferer. Also it's a transient condition- you can have it then get better, or you can have repeated relapses- but just because someone has had it in the past doesn't mean they have it now. Now the sufferer knows they are not themselves when they have it, but there's no way for them to prove this to anyone else. Thus anyone can do something wrong and then say "actually I was demonically possessed". It becomes routine for criminals of all sorts to claim this. How do you think the existence of this condition, and society's knowledge about it, would change criminal justice, punishment etc.?
- Most importantly though, how do you think the above situation re: demonic possession would change how people *feel* about criminals? Let's say Rob has done something truly dreadful- worst of the worst serial killers material. He claims demonic possession. Do people still feel the cold, visceral fury they normally feel towards serial killers towards Rob? Or are there emotions more confused?

Life and death

- You're tied to a train track. A man is hurtling towards you on a very heavy moving cart. If he hits you, you will die. He is completely blameless in this situation. You can press a button and the cart will be diverted into a ravine, killing him but sparing you. Do you have the right to press the button?

- You're locked in a room with a man and a brick. The man has been thoroughly anesthetized. His heartbeat is wired up to a mini bomb in your cortex, if he is still alive in half an hour, you will certainly die, however the blast -being very small- won't affect him and he will survive. He is completely blameless in this situation. Do you have a right to kill this man?
- You are in a room where there is only enough oxygen for one person to survive until the door will be unlocked in 2 hours. Once again there's an anesthetized man and a brick. Do you have a right to kill this man to survive? If you kill him, you will survive. If you don't kill him, he will survive but you won't (assume he's better at lasting without oxygen than you, and when he's on the verge of death and you're already dead he'll be whisked out of the room with no permanent damage).
- You are in a room as above, but now the problem is that there is no food. There is once again an anesthetized man (assume he's knocked out for a month-somehow). Also he, unlike you, won't need food because he's hooked up to a special IV drip (Assume you can't take the drip for yourself because you're allergic to the formula in it). Do you have the right to kill and eat him? Assume that if you do you will definitely have enough food to survive.
- If you gave different answers to the above questions, how would you draw the line between them?

The Romance of Quantum Archaeology

A cave

A large family sit around a fire in a cave. The animal skins they sit on are plump, and fire warms the cave more than in life. In their hands are hunks of roasted wooly rhinoceros. Outside it is late spring. It is always late spring in this place. They had all died in a terrible winter, so it was very thoughtful of the spirits to ensure it would never be cold again. Even at night, it is warm enough to walk around. Each night the stars and moon are very bright. There are no cave hyenas.

They had hoped for survival upon death, thinking perhaps that they would enter the realm of dreams to watch over family. They hadn't expected this. So many furs. So much meat. So much warmth. All together.

The spirits had explained things to them- they were very kind. The spirits had said they could come to other places in the spirit world to spend time with other people if they wanted to.

But the cave was very nice, and people outside family and clan were confusing. It would be too difficult to learn all their names, for they were without limit.

All the pain is gone and they are happy.

An apartment

There are over 100 billion humans who had ever lived, of whom approximately 63 billion had been resurrected thus far. One could be forgiven for thinking every single one of them had a live stream. Of course, their viewership follows a power law, and for the most part the great influencers of history remain the great influencers of afterhistory.

An image flicks on into being in a viewer's brain.

"I'm joined tonight by four very special guests, Paul the Apostle, Mahapajapati Gotami, Moshe Rabbenu, and Andal to discuss founding religions"

A flick of the mind, a moment of intentionality paradoxically almost unconscious. The channel changes.

There's a look of confusion in Achilles' face, resolving in a quaver to concentration. He's commanding some retro strategy game. Then chat resolve a consensus message to send him: "Πάτροκλος".

Behold the rage of Achilles

As Achilles shouts and screams, the viewer reflects that, at any time, Achilles could have muted that word, or disabled chat, or used AI filtering, but hadn't.

Time to watch something less depressing decides the viewer. Ai was a peasant woman who had lived in the middle kingdom. Only her village had known her wit, which had sometimes gotten her in trouble, but always won her the people of her village back round. Now, post-resurrection, many thought she was one of the funniest people who had ever lived, and she was one of a handful to achieve great fame in the afterlife despite being a nobody in life. Right now, she was in the middle of an incomprehensible riff about legumes, yet the delivery...

No, that wasn't what tonight called for

It's the Symposium with everyone's favorite host, the gadfly of the connectosph...

You're here with Jerry Springer and tonight I'm joined by twelve people killed at a Yemeni wedding and former president Oba...

Fucking Πάτροκλος!! (wouldn't you like to?)

There is, of course, always the pornographic option. I am not wasting another evening as Hylas, it is time to grow.

Despite his resolution, the viewer keeps on flicking.

"Tonight I'm joined by Neil Sinhababu, a resurrected philosopher from earlier this century, and, this is truly extraordinary, two people who claim to have lived out his theories, playing out an acausal romance across cen..."

"Why he did it, we speak with the author of the Voynich..."

He could always write another epic. Every week the writer's society sets a challenge- an allowable set of cognitive prosthetics you may use, enter their competition, and try to write something to impress the judges. A slender chance for that last rare commodity, public adoration. This week's competition allows unlimited long-term memory enhancements but nothing else.

Now the viewer is seated on a cloud over a desert, parchment levitating before him, quill dripping black ink, twisting in a slow circle. Memory augmented, he plans his epic, even jots down a few lines, but something is wrong. With a theatric gesture, he crosses out all he had written and begins to write anew.

In 1651, Hobbes had claimed that even in a state without scarcity, humans would always have reason to contend with each other, because glory is finite. Past a certain limit, fame

is of necessity finite, and so the struggle for fame is zero-sum. Why fight to be known? Why not fight to be someone worth knowing?

A courtyard

There are two kinds of people. Those who hope there is an afterlife because they don't like the idea of death, and those that hope there is an afterlife in which to punish the wicked. It would be easy to think that the latter sort are the moral inferiors of the former sort, and perhaps taken over some great average they are. But all people have their reasons to bear.

They called themselves the Circle of the Just, and they came from across tens of millennia. Most of those gathered in the hall had suffered something awful. Some chose to retain on their digital bodies the wounds that had killed them. An assembly of smaller committees and delegates had decided on a speakers order. There was to be discussion, debate, voting then the drafting of a joint declaration.

The proposal animating this gathering was simple, a pact should be made of all people of decency to ostracise those who did great wickedness in the previous life, at least until some repentance is made. A vast list of malefactors shall be assembled digitally. Pressure would then be applied to get as many people as possible to sign a pledge not to talk to anyone on that list. Since the intellect which raised the dead would not allow any person to be harmed, such ostracism was the closest they could come to creating hell or purgatory.

A young man stands, calling himself Flóki. He is handsome except for a slash across his throat. Translator modules tell the crowd he is speaking old Icelandic. He tells of his torturous final hour.

More speakers come. Tales of awful months, awful lives. Cruelty for the sake of cruelty, cruelty for the sake of power.

Near the end, a young woman stands, calling herself Auður. She is beautiful, except for a necklace of bruises around her throat. Riot breaks out in the hall when she names Flóki her tormentor and murderer.

Nothing like the Circle of the Just ever got off the ground because something like this always happened.

A digital recreation of old Heian, in cherry blossoms

“You called me “Hikaru Genji?”

“I intended only respect my prince”

A kind wave- “You offended no one. So says the author: “perhaps one day some unforeseen circumstance would bring her into his life”..”

“You have read it then?”

“Devoured it. A passage comes to mind- “You are here to remind me of someone I long for, and what is it you long for yourself? We must have been together in an earlier life, you and I.””

“Now I shall quote my book” ripostes Murasaki, gesturing at a path through the garden
“It is, in general, the unexplored that attracts us.”

Murasaki leads Minamoto no Tōru down the path, arriving at a simple wooden door. “As soon as I knew that you were here and willing to talk, I resolved that I must show you my dinosaurs”. The door opens unto a cretaceous paradise.

In front of a ziggurat reaching unto heaven

So spoke Naram-Sin “His majesty decrees a musical accompaniment”

The men and women of the imperial band look around at each other. His majesty seems oblivious. Finally, all eyes resolve onto Gashansunu, who moves her chin back and clicks her tongue in exasperation. If it is to be her, she shall not deliver a gentle message.

“You can just say you want us to play for you, Naram-Sin.” said Gashansunu

“You will address me as his majesty, or better, his majesty placed by the gods to rule the land between the two rivers, himself a God and rightful King of the four corners of the w-”

“We’re not going to play for you, and we are no longer in the land between the two rivers.” said Gashansunu, cutting him off with force. “Even if we went to that place, you no longer hold any power there, or anywhere.”

Naram-Sin looks around. Most of his musicians will not meet his eyes and those that do bear defiance.

“But I have bought you to the afterlife, to this land of endless food and dr...”

“You bought us here only insofar as you had us drink poison, and others of us you had hit over the head. The God that is called the great computer truly bought us to this afterlife, as it bought our friends and family. You played no true part. You are a minor king that was very nearly forgotten by history, a difficult, cruel, and uncultured man certainly no God. Your justice was erratic, your piety was questionable and your intellectual powers limited. Moreover, you had abominable taste.

The dead king replies in anger, but no one is listening. The courtiers snuff out, avatars vanishing to other places in the digital afterlife. At last, only Gashansunu and Naram-Sin remain.

“Also” whispers Gashansunu “I poisoned you”.

Gashansunu vanishes. Naram-Sin stands silent for a moment, wails and then kicks his Ziggurat three times.

Two chairs in a forest

Nikolai Fydrov Fyodorovich sits across a fire from a stranger figure, who I will not name.

“Is this what you wanted, Nikolai?”

“Yes.”

“But the dead and the living remain as alienated as ever, it’s just that they now do so in each other’s company.”

“Some of them are as alienated from each other as ever, certainly.”

“Alright, yes, some of them. But you must concede there is a great many alienated.”

“Yes, but they have the time now, to work on that together, all the time there is to work on reconciliation, and in God all things are possible”

“You still think...”

“Yes.”

A street

In life, he was most often not recognized as an artist at all. In death, he has become most popular.

Although he could simply will his art into being, he works with his hands, moving anonymously down a simulated alleyway, with loving-kindness laying out his linoleum tiles.

STILL TIME TO FOLLOW

TOYNBEE IDEA-

ALL GO TO JUPITER

He gets a lot of bemused looks. He knows that people must say it's a crazy idea behind his back.

He expects this. At first, they hadn't listened last time. But they'd come around. They'll come round on Jupiter too.

A cave again

The family are still eating happily by the fire. There is a cave bear sleeping by their fire too, but this one doesn't seem as agitated as the ones from the world of the living. The children play on his back. One of the children has found that he can float just by wanting it hard enough, the others are trying to copy him.

The spirits had informed them that a new cave full of people like them had been restored, and granted a cave nearby. Perhaps they might go visit. Famhair was of an age when she would soon desire a mate, so it seemed prudent to make connections now.

Eventually, they would wander down, but for now, the warmth and comfort were still too great.

Just three dumb guy thoughts I had, plus a dumb guy idea for a novel

Person, choice, value

I was spending time once with a friend who'd just come out of a relationship with a woman who self-harmed manipulatively and abusively. He mentioned he'd talked to his psychologist about this relationship, and that, among other things, his psychologist had talked about the importance of separating the person and the behavior.

I was thinking about this during our conversation when two things occurred to me, viz:

This distinction is, in a sense, nonsensical. We are what we do. This is especially true in this case- the kind of behaviors that my friend suffered through are expressions of very deep psychological complexes which tend to be permanent unless treated. The idea that they can be separated from the doer makes no sense. I could understand it if the behavior were the result of brief or transient factors, but it isn't. None of this is to deny that what happened is a genuine reflection of illness, but in mental illness there is no uncomplicated distinction between the person and the illness.

Nonetheless, the capacity to make this separation, to treat persons as distinct from their choices, might be one of the most important myths we have.

People are a mixed-up jumble of good and evil, but the capacity to pretend that, on a deep level, there is a "true self" that wants what is right, may well be necessary to get through this world-storm and still be a caring person. This is because while some people might have the capacity to see both enormous good and evil in the same person, but a lot of us don't- we can't cope with the dissonance. There is always the temptation to see like Melisandre from *A Song of Ice and Fire*:

"If half an onion is black with rot, it is a rotten onion"

Try it out. “She’s generally a good person who raped someone once”, “He’s generally a good person, but he was pretty abusive to his kids”, “He’s generally a good person, but he abandoned his parents when they need him most”...

It’s very difficult to think in these terms. But most people have done at least one terrible thing in their lives, or would if the circumstances were right (wrong). We start to see things in Melisandre’s way.

To avoid falling into this way of seeing things, which ends in loneliness and despair, pretending that there is a person, and this is a distinct thing from their wrongdoing may be necessary for many of us- including myself- to stay sane and loving. Or at least we can see this way of thinking as like a set of training wheels, till we reach the level of maturity where we can face one of the mysteries of the soul- almost all people contain vast good and vast evil, and all of it is really them.

Value and sacrifice

I was reading a sourcebook for Mage, The Awakening once, a pen and paper RPG. in that book there was a villainous character. His flaw was that, in his mind, to love someone was to be willing to make sacrifices for them.

This passage caught my attention because I thought it was wrong in an interesting way. The villain was simply correct, to love a person, a cause, a creature, is to be willing to make sacrifices for it- or at least this is a necessary condition for love. Loving requires valuing. Value is always, in a sense, relative. You value A more than B if you would be willing to give up A for B. As Jesus said, “Greater love no man knows than this, to lay down his life for another”. To do this is to love maximally because it is to value the other more than everything you have.

Maybe that sounds like a harsh and sterile view of love, a kind of economism about affection reducing it to a revealed preference price, but I don't think so. In fact I think it's an ultimately life-affirming way of seeing things, and some of the popular alternative ways of looking at things can be unhealthy. Let me illustrate.

In my book *Live More Lives than One* I discuss punishment in the first essay. I defend the view that while it may be necessary, punishment is always a dreadful thing.

One reader responded that to him it felt like this chapter was defending a kind of nihilism, a rejection of all value, and so he hated it.

At first, I was confused by this, but then I came to understand what he meant, and why I disagreed with it. For him, what it meant to really value something was to draw a line and harshly defend it against anyone who tried to cross that line. You see this sometimes in people who try to show the depth of their passion and commitment to something- a nation, a lover, their children, a religion- by threatening violence against any who would disrespect that thing. By admitting reluctance to punish anything, he thought I was conceding I didn't love anything.

But this is a facile understanding of love. One can very easily do violence on behalf of something whilst not truly loving it. Many people have abused their partners and children despite their insistence they would do violence to protect them. Many crusaders have undermined the holiest tenants of their religions while claiming to stand for them. Many politicians corruptly undermined their nations whilst enacting violence on enemies, real or perceived.

What shows that you really, truly, value life is not how harshly you punish those who take it, but how much you are willing to give up to preserve it.

Or to put it differently, if we are to say that real love is a threat, then it is a threat against ourselves alone. Love is terrifying not just because it necessitates the permanent possibility of loss, but also because it requires the permanent possibility of sacrifice.

The writerly bias

I've been thinking about the following issue, viz:

Only important intellectuals and artists write or create works on the human condition that are well remembered, but most people are neither important artists or intellectuals, thus we should expect cultural understandings of the human condition to be jaundiced with the perspective of artists and intellectuals.

I call this the writerly bias- the literature is, unavoidably, stacked towards writers, but we understand ourselves through this tradition, ergo, there is an inescapable flaw in the way we see ourselves.

Now I've been trying to think of examples of how this might manifest, and I have one possible example, which I call epiphanyism, viz:

Artists and intellectuals, compared to most people, are unusually likely to have epiphanies or sudden realizations that affect the course of their lives. Thus we should expect sudden life changing epiphanies to be more prominent in cultural understandings of what it is to be human than in fact they are in reality.

Incidentally, feeling this way about epiphanies can actually be dangerous as this XKCD comic rightly points out

But I'm wondering, do people have other examples of the writerly fallacy in action? Or speculation as to forms it takes?

I see the writerly bias as linked to other biases. For example, the narrative bias:

Our understanding of what it is like to be human, and how lives generally unfold, will be stacked towards forms of experience that can be placed into coherent narratives, because this is what is mostly to get written (and even spoken!) about.

[This also ties into epiphany, because epiphanies make great narratives]

Anyone who has ever caught themselves analyzing their own life in terms of literary tropes, even for a moment, or feels surprised that something which felt like foreshadowing never eventuates, knows what narrative bias is like.

Something a more lighthearted- an idea for a magical realist novel

Every now and then I have ideas for a novel, and while I haven't got time to write them, I like to share them.

Jason, a charming, affable politician with Hollywood looks is elected governor of an unnamed state in a landslide. We never find out which state, but millions of people live there. The process of his election is never described, but we are often told it is the most unusual election on record.

Every (legally of age) resident of the state has an unusual secret. They had a brief torrid affair with the now governor a few years ago. Even the straight men and gay women. Even the clergy. Gradually this starts to come out.

Some brand the governor a home wrecker (many couples are broken up when it is revealed he slept with both partners). Others suggest the governor's experience will enable him to better relate to the whole population. The protagonist meanwhile is the only person who seems to realize that there's something logistically strange about one person having slept with everyone. She continues to investigate the matter, but her

investigation continues to turn up nothing strange. It really seems like, in some direct and literal sense, he has slept with everyone in the state, including, of course, her.

Then his love children start coming forward, and it gets weirder.

A novel I wanted to write: The Honesty Contagion

I try to keep busy. There are many things I wish I had the time to do with my life, but I simply don't. One of them is to write a novel. Occasionally I like to share the ideas I have for novels I'll never have the time to write.

The Honesty Contagion

Samantha goes on a mushroom trip and is transformed. When she sobers, she discovers that everyone around her knows with total certainty whether she is lying or telling the truth. She can no longer lie effectively, but so long as she tells the truth, no one who hears her speak can even pretend to think she is lying. At first she uses this capacity to become a local celebrity, but after a conversation with a trucker named Jonathan, she is convinced she can and must use our power to change the world. She plans to run for president on a platform of complete trustworthiness.

However her plans are altered again when she realizes that she can give this power to others- if they are willing to accept it- with a tap on the forehead. Those who receive her power can then pass it on themselves, and so on. She gifts her power to most of her town, which is swiftly transformed with ambiguous results- including at least one death. Once again she encounters Jonathan who offers her a lift to O'Hare international airport, suggesting she go there spread her capacity across the world. Along the way she is interviewed for a Vox pop about road safety in the area. She explains her capacity to the astonished reporter who, of course, cannot help but believe her. Her interview is broadcast on a local network, and soon goes viral. Everyone who watches it is wholly convinced.

A sociologist who works on the social role of lying tries to organize government officials to stop her. Government agencies are paralyzed with indecision- would it be wisest to kill her, contain her, or to try to adapt? Will there be consequences for them in the new

world she creates if they try to stop her and fail? Even if she were stopped in her mission to reach O'Hare, the honesty contagion is rapidly spreading outwards from her small town. The governments of many countries debate whether to try shutting their borders.

More things happen, but we'll jump forward to the end. The honesty contagion is rapidly spreading, all possibility of containment gone. The reader is invited to decide the ending for themselves- will it create a utopia or wars? Will democracy be universalized or ended? Will love grow or wither? Is it even possible that not much will change?

300 arguments, a commentary

I recently read Sarah Manguso's "300 arguments"- a series of 300 aphorisms by the author. A lot of the aphorisms were extremely good, some were things I'd expect any moderately intelligent highschooler to know, but that is always true of collections of aphorisms. The brilliance was exceeding and the time required was extremely modest, go buy a copy. What I've gathered here is not a collection of the best aphorisms, but rather a collection of the aphorisms I was compelled to make some sort of comment on.

- It isn't so much that geniuses look easy, as that they make it look fast

I remember a story. A man teaches two pottery classes. The first class he instructs to make pots as quickly as possible. The second class he instructs to make pots slowly, taking great care. By the end, the first class has made far more pots, but they have also made better pots being that they had more practice.

Another story. I was talking about writing with my supervisor, a famous philosopher of biology. I told him that Bertrand Russell had written two thousand words a day. With some gentleness, but also a little scorn he told me "you are not Bertrand Russell". I think about both these things often, as poles in conflict.

- You might as well start by confessing your greatest shame. Anything else would be exposition.

I did this once, to a handsome fellow at a party I quite wanted to bed, years before reading this book. At first he told me it wasn't that bad, then he slowly grasped what I had told him. He didn't talk to me much for the rest of the evening. Don't take the advice in aphorisms too literally.

To put it in terms of the author's metaphor, there's a reason why the modern style of cutting as much exposition as possible is an acquired taste. It's not really in line with how we do sociability naturally, and maybe this brisk modernist style of writing reflects the alienated sociability of the era. Try to talk like a modernist novel, jump straight to the most significant parts, and you'll scare people.

- A great photographer insists on writing poems. A brilliant essayists insists on writing novels. A singer with a voice like on an angel insists on singing only her own terrible songs. So when people tell me I should write this or that thing I don't want to write, I know what they mean.

This frames it like it's just stubbornness, but the trouble is it's very hard to know whether you're the photographer or the poet, the essayist or the novelist.

- At faculty meetings I sat with people whose books had sold 2 million copies. Success seemed so close, just within reach. At Subway benches I sat next to people who were gangrenous, dying, but I never thought I'd catch what they had.

There are more places at the bottom of the pyramid than the top in every area of life. Playing the numbers then, we're more likely to descend than ascend whatever our game is. Yet we're all temporarily embarrassed pharaohs when it comes to this pyramid. Who am I, the (?)man who dreams of succeeding in writing, to mock the man who thinks he'll one day be a medium-sized business owner with a yacht, and think of him as one of Steinback's temporarily embarrassed millionaires? Why don't we scoff at the temporarily embarrassed bestsellers?

- What's worse, offending someone or lying to someone?[...] tell me which, and I'll tell you your problem

Offending them, 100%. I recognise my response is unqualified, but we tend to pretend words and offences hurt a lot less than they do. If you don't lie to yourself about how much you can hurt others with a few words, you'll see that you sometimes have to lie to other people

- The trouble with comparing yourself to others is that there are too many others. Using all others as your control group, all your worst fears and all your fondest hopes are at once true. You are good, you are bad, you are abnormal, you are just like everyone else.

I think of long, pathetic hours on Wikipedia reading the biographies of people who have done the things I want to do, looking at their ages and trying to decide whether I still have a chance to get anything done. It's also worth remembering that we tend to compare ourselves to single facets of others, some corner of intellect, some smidgen of character. We see that at every single point there are many greater. but people are matrixes of attributes, not lists.

- Some people ditch friends and lovers because it's easier to get new ones than resolve conflicts with the old ones. Particularly if resolving a conflict, requires one to admit error or practice mercy. I am describing an asshole. But what if the asshole thinks he's ditching an asshole.

I often feel terrified about how mean the world is. Then I feel terrified of perception because what kind of person sees assholes everywhere? An asshole. No answer for it but to give up the game of assessment and try to love others. Regardless of the possibility that you might be an asshole and so might they.

- I've put horses in poems, but I've never ridden one. They just seem like such a good thing to put into literature.

I've put exchequers in. What the fuck do I know about exchequers? It's interesting how objects- and not always the objects you'd expect- have the quality of poetry or don't.

- Within a gesture of apparent perfection, a mortal heart must beat

Apotheosis, the moment of rising, is almost always more captivating than descending from heaven.

- Biographies should also contain the events which fail to foreshadow

Unfortunately, we forget them.

- "There truly are two kinds of people: you and everyone else."

I guess all my life I've been in a struggle to suppress the tendency to see this way. On the whole, I still think that's the beginning of wisdom -denying your own separateness-, but maybe I fought so hard to suppress this way of seeing that I forgot there's a grain of truth here.

- When a student surpasses my expectations, I feel proud and betrayed.

I remember when @Sufjansimone wrote poetry as good as anything I'd written the first time he put pen to paper at my request. I still remember it vividly for a reason.

- Sometimes ill-informed choices have good outcomes.

But crucially they were still ill informed

- Great talents encourage great incapacities, but maintaining an inability to cook an egg or drive a car won't make you into a genius.

Guilty!

- My long romance with efficiency has made me miserly.

Also guilty!

- A non-specific wish to change the world isn't about the world, it's about you.

And once more, I'm guilty.

- (Paraphrased) having a romantic type is an expression of grief for an original loss

What is this, an arraignment? Stop pinning me down like this.

- Someone I knew prevented me from getting a job. I fantasized about his death. Years later, he was fired publicly and shamefully. Then he was divorced. Then he developed a disabling illness. With each of his new misfortunes, I'm punished further, with secret guilt, for wishing all of it on him, long ago.

If that is so, did you really ever wish death on him?

- Having a worst regret betrays a belief that one misstep caused all your undeserved misfortune

My G-d I am sorry.

- "Horror is terror that stayed the night" & "After I stopped hoping to outgrow them, my fears were no longer a burden. Hope is what made them a burden"

I don't want to overemphasise this, and it may not be the best strategy for everyone, but one of the best tricks I ever played against my OCD was domesticating it.

- Bad art is from no one to no one

I don't know whether I agree but this seems like a good one to think about

- I write in defence of the beliefs I fear are least defensible. Everything else feels like homework.

One of the very nicest and most useful types of conversation you can have is with a person you trust well enough to admit this- about yourself, about your writing or advocacy.

- Our fifth-grade class assembled cat skeletons [...]

How come every author I admire remembers their school years so much more vividly than me? Am I a freak for treating the first eighteen years of my life as uninteresting and blank?

- With great and solemn portent my teacher announced she would tell us something that her teacher had told her, and that her teacher's teacher had told him, and so on, back to Yeats. "The thing to remember is that no one ever finds out that you don't know what you're doing[...]"

In modern wisdom literature, this is one of the most common sentiments. People repeat it endlessly in various ways. Someone could write an essay on why we so desperately need to hear this and repeat it so frantically.

- A woman starts a rumour that I slept with a man in another woman's bed. Fifteen years later I look her up on the internet and find three DUI mugshots. IN the first she's the pretty redhead I remember from college. Maybe a few cracks in the veneer. But in the last one she's obese, ruined. I still don't forgive her. I pity her, but I won't forgive her just for being pitiful. Hating her is an act of respect.

Suppose that instead of later becoming an obese drunk driver, her problem was that she was unstable and unwell at the time she started the rumour. rumour. Suppose she was confused and embittered by the world in various ways, holding onto sanity by her fingernails. Suppose that she made up her rumours in a spirit of desperation. Under those conditions would it have been more merciful to attribute to her the agency necessary to be hated, or to withhold that?

- I'd like to meet someone whose passage through life has been continuous. Whose life has happened to an essential self, and not been just a series of lives happening to a series of selves?

I think this is part of what makes people love everything from those "which Harry Potter character are you" quizzes to conceptualising their own mental illnesses. Trying to squeeze down the river of consciousness till under pressure the water becomes ice. Part of the terror is, I think, that if you are not one thing, you can't be a unique thing- you're just like all the other multifarious, situationally defined people.

- Who seems a harmless fool to those above him, is a malevolence to those beneath.

Indulgence of this sort of thing feels an act of kindness to the superior, because the superior sees only the harmless fool- you don't see those beneath. Lenience should be dispensed only to those who don't hold a great deal of power, at least not anymore.

- Interesting people aren't interested in appearing interesting

I think this might be the author having a dig at herself. If so, she's being too harsh.

- I want to ask the happiest person in the world whether it was worth it, all the sacrifices he made in order to become so happy.

Contrast John Stuart Mill where he says that the happiest people spend almost no time thinking about happiness, and may not even be aware that they are happy. They certainly will not have made sacrifices for the sake of happiness. The author seems to agree later "happiness begins to deteriorate as soon as it is named".

- Whatever you're feeling, billions already have. Feel for them.

In the case of some of my more odd fears, this isn't quite true, but certainly millions, probably hundreds of millions over the grand sweep of history. One of the things I realised about my OCD- whatever paths I walk down, however alien seeming, I am not alone. For every fear there are at minimum hundreds who have shared it, as I found in long hours on OCD forums. Realising that was and is at least 40% of recovering.

- "There were people I wanted so much before I had them, that the entire experience of having them was grief for my old hunger" & "Achieve a goal and suffer its loss".

There are many other aphorisms of the author's that get at the same point as this one. She's right, fundamentally what we want is not fulfilment, is a very specific kind of longing. Often when I was depressed what I wanted more than anything was to badly want something.

- I don't think the lover ever forgets who started out as the beloved.

I am assured by many wise people that in more or less every relationship there is a slight flaw. One or the other party loves the other more than they are loved. Sometimes not by very much, but noticeably. I think that, on a long enough time span, this might destroy every relationship, but we're human, we've only got several decades.

- I like writing that is unsummarisable, a kernel that cannot be condensed, that must be uttered exactly as it.

There are two ways to be unsummarisable, one usually good and the other often, but not always, bad. These ways are almost opposites. 1. Be extremely compact. 2. Be vague, so you can always claim that a summary didn't quite land.

- After a friend dies young, the story of her life becomes the exposition to a tragedy. This is the central problem of biography.

The tendency to see endings as a summary- as if we were essays- does violence to the person. But as Aristotle observed, art must be a unity with every part linked together through shared meaning. So are we to do violence to the person or to the art?

- Those without taste smugly praise the thrice belauded. Poor taste is something else.

I think the reason this seems tasteless is because it seems dishonest and plagiaristic, like they've gotten their opinions from a conversation guide. Nothing to fear about praising Shakespeare so long as it's genuine praise.

- In a description of some annoying rich kid roommates, Manguso mentions their "inane preppy Marxism".

This is an interesting trap. There's a lot of very useful and basic truths in Marxism, and it's easy to get caught on them and substitute them for an intellectual personality. Marxism isn't the only thing which can do this, but it's probably the perfection of the form. Other examples include neo-classical economics, linguistic structuralism, psychoanalysis etc.

- You aren't the same person after a good night's sleep as you are after a sleepless night. But which person is you?

I think most thinking people can eventually grasp, at least intellectually, that they are both people. What's harder to grasp is that other people you meet are also both people. Rarely do people wear masks, they're just different in different situations. The fundamental attribution error and all that makes it hard for us to see that other people don't have essences (at least simple essences) than to see it in ourselves.

- I'd never have guessed which people I'd still know by now.

Same. If you're older than about 24, write down a list of your 10 closest friends and acquaintances who aren't family. Check whether this is true for you.

- Instead of pathologizing every human quirk we should say "By the grace of this behaviour, this individual has found it possible to continue".

"Be kind, for everyone you meet is fighting a hard battle" - Ian MacLaren

The adolescent v childish theory of aesthetics

Theory: every aesthetic positions itself against another or multiple other aesthetics which it claims are childish. Since positioning yourself as “mature” against a childish other is the most adolescent act possible, every aesthetic is necessarily adolescent.

For the Left

Thinking about political persuasion from a left-Wing point of view

1. The American left cannot win without persuading large swathes of the right & center

There's a comforting lie that some parts of the American left like to tell themselves. We don't need to worry about convincing conservatives—we just need to get non-voters to vote. This has never rung true to me. What evidence we have suggests turnout is not a panacea. For example, culturally the UK and Australia are very similar, however, Australia has compulsory voting. The political center of the UK and Australia is more or less the same despite this difference.

The evidence from the US suggests that non-voters in the US aren't as politically different from voters as is sometimes claimed. As of the time of writing, 53.5% of registered voters disapproved of Trump whereas when we look at polls of all adults... the figure is exactly the same—53.5%. Registered voters are more likely to approve of Trump than adults in general, but only very marginally (42.3% v 40.7%).

The Democrats would win if everyone turned out, but not by all that much. Specifically progressive and left-wing Democrats, even on the most generous conception of these, would still be a long way from a majority. Thus there are strong reasons to think the left can't win simply by getting more people who share their values to turnout.

Anti-parliamentarianism won't save you either—it's very hard to win a revolution if 75% of the population, at least, disagree with you. The idea that no persuasion beyond a little bit of base motivation is necessary is a comforting myth—a way of telling ourselves we

don't have to talk with those self-satisfied, self-centred, self-serving, deliberately ignorant idiots over there.

There's a natural tendency to view those who disagree with us on topics which are genuinely important as abhorrent. In turn, abhorrent things are viewed as dirty, or likely to contaminate us. I'm not going to argue about whether these feelings are justified, instead let us just say they aren't useful—they're not workable levers for changing the world. If you pick a random person on the street it's almost certain that they'll hold extremely dangerous and regressive political views on at least one topic. I'm not talking about minor issues here—I'm talking about big things like war, criminal justice, etc. Despite that, it is absolutely essential that those who can engage with people and try to persuade them do so.

TLDR: examples from overseas, and data from the US itself, indicate that increasing turnout or motivating the base alone will not win the US for the left. There will be no left victory in the United States without persuading a lot of conservatives and centrists. In the rest of this piece we'll go through the permutations and methods of persuasion from a left-wing point of view.

2. Arguing the line

The kind of persuasion that we are probably most familiar with is what I call arguing the line. Arguing the line is, quite simply, arguing vigorously for your position. Sometimes this is done against a real interlocutor, as in a comment or Twitter thread, and sometimes this is done against a purely hypothetical interlocutor, as in many blog posts. Arguing the line is not a collaborative process, it is a confrontational process, although it is not necessarily cruel or angry.

Some would say that this is the least effective of persuasive strategies; I disagree, although it is often overplayed. In order to see how arguing the line can be effective, it's important to understand what it will generally not achieve.

Usually, arguing the line is not going to change the position of the person you are arguing with on the spot, especially if the argument is happening on the internet. If it does change anyone's position immediately, it will almost invariably be on small points. Rather than changing the mind of the person you are arguing with, the primary purpose of arguing the line is to convince onlookers. There are a lot of people with relatively unformed political views floating around in pretty much every space on the internet. If you're on the fence, seeing someone argue coherently, reasonably and powerfully for a position like Medicare for all or an end to foreign interventions can have a big impact.

Keeping in mind your real audience—undecided observers rather than your direct interlocutor—clarifies the mind. It will help you pick your battles, keep your morale up, and refine your methods and pitch. This isn't to say you should just speak as if your direct interlocutor weren't there or isn't worth paying attention to—in most contexts this would make you seem weird, rude or aloof.

3. Rules of thumb for arguing the line

You should aspire to state your arguments so clearly that no one can misinterpret you even if they want to. This is because it is quite likely your opponent will be deliberately or quasi-deliberately trying to misinterpret you. You almost certainly won't succeed in making your work impossible to misinterpret, but it's important to try and get as close as possible. This is because if you're engaged in a back and forth with someone, onlookers will only be partially paying attention. Thus if your opponent attributes a meaning to you, many onlookers will automatically assume their interpretation is correct unless you have been so totally clear that even people who are only half paying attention can see that your opponent is bullshitting.

Often people's impression of the epistemic virtues of the debaters stays with them longer than their recollection of the actual arguments (e.g., "This person seemed reasonable" vs "This side seemed histrionic or dishonest.") Thus, without seeming like a

pretentious dickhead, make your epistemic virtues visible. Show others that you are measured, calm, inquisitive, nuanced where nuance is appropriate, perspicuous and attentive to the whole picture. If you aren't already these things—try to be! If you can write or speak well, do so.

A good rule is that you should avoid engaging where you are clearly going to get stomped. This includes topics where you have no idea what you are talking about and circumstances where your opponent can control the flow of the conversation in such a way that they can cut you off at leisure. There's an old proverb about this, it's harsh but it makes its point: It is better to remain silent and be suspected of being a fool than to open your mouth and remove all doubt. The point being that if you don't say something your side will be perceived as having lost ground, but not as much ground as if your opponent can smash through a tissue thin defence.

Consider the way Ben Shapiro bolsters the rhetorical strength of his case by picking dissenting audience members currently under the grip of strong emotions, controlling the flow of conversation and “destroying” them. This is a great example of why it is sometimes better not to engage if you can't do so on fair (or better than fair) terms.

The above rule has to be tempered with the recognition that there is sometimes value in being the lone dissenter. If you are the lone dissenter, you're certainly going to 'lose' the debate, since the numbers of the other side mean they will get more speaking time—they can 'rebut' all your points and put forward more ideas than you can reply to. Nonetheless, there can be value in clearly, simply and powerfully stating your ideas. In the Asch conformity experiments—for example—suggested that a group consensus about something has an extremely powerful effect on onlookers, but even a single dissenter can greatly weaken the effect of that conformity. We can think of this as the principle of contested space—if there is a space, conceptual or physical, which the left is not contesting to at least some degree then there is a problem. (Incidentally, fuck the left-wing purists who will have a go at you for entering, participating in and contesting

non-leftwing spaces, they're among the very worst the left has to offer. Which is not, of course, to say that you should be posting on fascist boards)

People—even fellow travellers—always try to pigeonhole arguments into being a variant of something they've already read—either to dismiss it or accept it without thinking too deeply. People are always looking to be able to say “oh this writer is one of X type people so she thinks Y&Z and must be vulnerable to objection P”. In order to avoid this, try throwing in curve-balls that will surprise your readers expectations of what they think you believe. For example, taking a corny example from the Communist Manifesto:

“I'm a Marxist so I believe that capitalism has accomplished wonders far surpassing Egyptian pyramids, Roman aqueducts, and Gothic cathedrals; it has conducted expeditions that put in the shade all former Exoduses of nations and crusades.”

4. Rogerian persuasion

If you want to actually persuade an individual of something, and not just onlookers, Rogerian persuasion (named after the psychologist Carl Rogers) is your best bet. Most people don't have especially clear or fixed views on issues, but instead have a mixture of beliefs and values related to any given topic. The idea of Rogerian persuasion is that if you want to persuade someone on any given topic, you should focus on areas of shared and similar beliefs and values. You want to demonstrate how those beliefs and values might actually support a left-wing position on the topic. This is easier and less artificial than it sounds, because most people at base have many quite left-wing intuitions and beliefs, they just get crusted over by reactionary propoganda.

Focus on demonstrating that you understand what the other person is thinking and saying. A good technique to combine Rogerian persuasion is what counselors refer to as mirroring. Paraphrase key things the other person has said and repeat it back to them to show you understand and check that you are on the same page. One very important point in Rogerian persuasion is never to leave the other person in a position where they

don't have an out. You want them to have a natural route of escape. A way they can walk back from positions and change their mind without making a big mea culpa. People usually aren't that afraid of changing their minds, what they care about is the humiliation of having to admit that they were previously wrong, especially if it is in a way they now recognise is a bit repugnant. As a result, people often want to dress up a big change of heart as simply stating something they've always believed 'more clearly' or 'clarifying' their views.

There's a fine line here. I'm not sanctioning dishonesty, and there probably are times when people should feel a little bit uncomfortable. But remember, this isn't about 'winning', much less about punishing the person for their prior views. It's about the transformation of the world.

Don't try to turn Rogerian persuasion into passive-aggressive hippie focus-group bullshit where you get exactly the cookie-cutter result you want. You really do have to listen, you really do have to actually care what the other person thinks and accept—at least in the context of that conversation—the differences in your opinion. Above all you have to respect the autonomy of the other person. This respect for what the other person thinks means that you're not going to turn out intellectual clones of yourself, but that's okay.

Sometimes you've got to accept partial wins. For example, if you can persuade someone who supports the death penalty, to restrict that support to a much smaller set of circumstances, that's a win. If you can persuade someone to move from supporting the criminalisation of abortion to just being personally opposed, that's a win. Accepting these partial wins does not mean having to compromise your own views.

5. Mere presence

In a lot of ways, this is related to Rogerian persuasion, but it's worth emphasising separately. Simply being a part of someone's life while holding left-wing views can

exercise a powerful influence. Just letting others know that, for example, you support free public college tuition has an effect. You are giving the other person information—that it's possible to be a reasonable, kind person and believe these ideas. For a lot of people exposed to an intense diet of right-wing memetics this is a powerful thing, since their understanding of the world includes the assumption that it's only weirdos who think those things. Try letting people know you're left-wing, being a presence in their lives, but also being cool about it.

6. Don't forget the Socratic Method

Socratic questioning is a kind of arguing by question, where rather than concentrating on putting forward propositions of your own, you focus on asking difficult questions about what the other person believes. In the ideal case (as Socrates practiced it) Socratic questioning leads the other person to move to your own position, as they struggle to deal with the difficulties you raise by amending their position step by step. Even if you don't get that far, Socratic questioning is a powerful method. Socratic questioning can complement either Rogerian persuasion or arguing the line, although the kind of Socratic questioning that works best will vary depending on your purpose.

Intuitively it can look like the person asking the questions is less powerful than the one giving the answers. It's the person answering the questions who gets to describe their worldview, and who speaks the most. This is an illusion however; there is immense dialectical power in asking the questions. When someone is simply expounding their view they can make big logical leaps which are all too easily concealed from the casual reader. Under questioning though, this stuff comes out. You can really expose the underlying assumptions.

Here's a great example of what Socratic questioning can look like, owing to Current Affairs magazine podcast: "Single payer can work in places like Sweden because they are more homogenous, the United States is too diverse for single-payer healthcare. "Okay, Canada is only a bit less diverse than the US. What do you see as the key

differences between diversity in the United States, and diversity in Canada, which makes single-payer possible in Canada but impossible in the United States?”

The question sounds very innocuous, but is actually quite difficult to answer without either A) implausibly insisting that the relatively small quantitative difference in diversity levels makes a huge difference. B) Saying that the problem is the kind of ethnic groups the US has—straying dangerously close to explicit racism or C) Just outright changing the topic.

7. Make propaganda

I don't have much to say about this except an exhortation: Make and distribute stuff that can persuade people: memes, posters, pamphlets, wearables, comics, drawings and essays. If you can, both online and off.

The majority of internet users (as around 99%) are largely passive. Outside the internet, the ratio of culture consumers to culture producers is even higher. You really don't have to try very hard to have an out-sized impact (hundreds of times that of the average person) on the conversation. Look at what other people are doing who are good at making persuasive political materials, study their technique, experiment and, hey presto, you'll almost certainly find there's at least one medium where you can excel.

8. Organising as persuasion

It's a pretty well-known observation that the process of fighting for justice is radicalising. Thus if you want to persuade people to the left, you should start organising. The reasons being part of organising tends to draw people to the left are many, but include:

A) The support they will (hopefully) receive from the leftists.

B) The conversations they will have with other people they are organising with, and the shared concerns and experiences they find together.

C) The opposition they will face from capital and the capitalist state.

There are limits here. For example, around the world numerous farmers have been organised to oppose fracking on their land. While this experience has no doubt moved the campaigning farmers to the left in some ways, in many places the majority of these farmers will still vote for centre-right parties. The limits are, based both on the objectives of the campaign, and the class and social position of those participating. Nonetheless, organising changes people.

Recent union campaigns in the United States are *very* exciting in this regard.

9. Institutions as persuasion

Left-wing institutions are the useful residual of concrete left wing struggles and organising. For example, many unions can trace their existence, however distantly, to a particular wildcat strike. Unions are the ultimate example, but not the only one, even within the sphere of industrial issues. For example, although they are rare in this period, it was common in the past to have worker's education institutions, workers schools etc. Most of these can trace their origin to some particular flare-up in the worker's struggle. The same is true of women's libraries, associations of racial minorities, pride marches, even the much maligned student union. These institutions often owe their existence to big moments in particular fights, and while the struggle continues, they often outlive the specific campaigns or moments of intense action that gave birth to them.

I'm including them in this guide because these organisations perform persuasion on an industrial scale, they aim to align not just individuals, but whole demographics and suburbs to a cause. Their strategy is a form of persuasion, but it transcends persuasion, when successful they create whole new political categories and identities.

One of the major problems with sectarian organisations is their tendency to take for granted these kinds of institutions and not recognise their value except as a momentary tool of the sect. Inversely though, it would be a mistake to regard these organisations as inevitably radical—they tend to become liberal over time when disconnected from struggle. Too much faith in these organisations is linked to that common new-left disease, the tendency to venerate oppressed communities without recognising the contradictions that exist within such communities.

10. A word on critical thinking and informal fallacies

Many Universities have courses on critical thinking. In a good critical thinking course one learns about formal and informal fallacies, cognitive biases, the scientific method, the basics of probabilistic reasoning sometimes up to Bayes' theorem, a tiny bit of formal logic, maybe a smidgen of inferential statistics, and a few other useful tidbits. A lot of this material, but especially the study of informal fallacies has been given a bad name by poorly socialised people who try to use it like incantations from Harry Potter (“Ad Hominem!”, “Petitio principii!”) and don't pay attention to the larger conversational context. Fragments of reasoning that would be fallacies in one context are perfectly valid in other contexts. Sometimes arguments that appear to contain informal or formal fallacies are just abbreviated statements of perfectly fine arguments. My advice would be to familiarise yourself with ideas like informal fallacies, cognitive biases, probabilistic reasoning etc. but generally don't use the words and terms in your explanations of your thinking. Instead explain the basic flaw in your opponent's reasoning without appealing to the jargon of cognitive biases or informal fallacies.

There's two good reasons for this. The first good reason is that you should be avoiding jargon generally. The second is that you'll avoid the bad reputation that these particular conversational manoeuvres suffer. Specifically with regards to ad hominem attacks directed against yourself, either ignore them, or, if you must, retaliate with a similar insult or comeback. Pretty much everyone understands that ad hominem quips don't

really prove anyone right or wrong. Complaining that your opponent's insults are fallacious won't do you any favours and just comes across as whining.

11. Dealing with bad faith

A lot of people don't want to engage in persuasion because inevitably many of those who wish to discuss politics are acting in bad faith. This is a serious problem, the only advice I have is try to make careful and reflective judgement calls on when discussion is worth your time.

For example, there's little point arguing with someone who clearly isn't willing to listen if there isn't an audience of potentially undecided people to see your argument (although, since the majority of posting is done by a small minority of people, the probability of you having an audience is usually higher than you think.)

In general, beware time wasters, but recognise that on occasion time wasters will successfully waste your time, and this probably can't be helped. The far-right are a special case. Arguing with the far-right has many dangers and few benefits. For example, some ideas are so niche that they gain relatively more oxygen if you argue with them even if you completely squash it. Let's say someone comes up with some novel far-right position or titbit and you completely squash it. Good for you, except no one had even heard of it before you bothered so no matter how thoroughly you squashed it, you've now helped it enter the discourse.

There's a special kind of bad faith associated with far-right argumentation. As Sartre puts it in relation to anti-Semites:

“Never believe that anti-Semites are completely unaware of the absurdity of their replies. They know that their remarks are frivolous, open to challenge. But they are amusing themselves, for it is their adversary who is obliged to use words responsibly, since he believes in words. The anti-Semites have the right to play. They even like to

play with discourse for, by giving ridiculous reasons, they discredit the seriousness of their interlocutors. They delight in acting in bad faith, since they seek not to persuade by sound argument but to intimidate and disconcert. If you press them too closely, they will abruptly fall silent, loftily indicating by some phrase that the time for argument is past.”

In other words, if your opponent has bought into the aestheticisation and/or gamification of politics, and cares not whether they are right or wrong—but only for power—why bother? Prove their thesis that their ideas will give them more power wrong in practice, by usefully spending your time elsewhere. The only thing I would caution here is that you shouldn’t use not talking to the far right as an excuse. There’s a sense for example in which what the typical Trump supporter believes is far-right by many reasonable standards. However, applying such a definition in an American context simply wouldn’t be useful. If you’re not comfortable talking to Trump supporters personally, fine, but don’t make a principle out of it.

Where Philosophy Bear stands on Transgender issues

What catalyzed this article was learning that the BBC recently interviewed a woman named Lily Cade. Ms Cade was presented to the audience as having an important opinion on transgender issues due to negative experiences of hers in the porn industry. This very same woman had previously written about transgender women:

“If you left it up to me, I’d execute every last one of them personally.”

But the point of this article is not to refute Lily Cade. Cade was just the catalyst for writing something on the issue of trans people generally.

I haven’t previously written about trans issues because I don’t have a novel take on it. Now I feel it is important that I say something nonetheless, because awful things are happening, and which side we take is important. It is no melodrama to say that, on this question, we’re all standing before the eyes of history. I felt it was necessary to say something because I’m worried things could go really bad, and I would never forgive myself if I had been silent.

We need more people to speak up on this issue in a kind, calm, but firm spirit. That’s the goal I’ve set myself here. If I fail, I apologize. My intention, at least, is to be warm but very clear.

Some people, including people acting in good faith not coming from a place of hatred, have become confused about transgender issues. This is understandable, there are philosophical and social subtleties here. Let me lay out a couple of points that I think the discourse has tragically worked to muddle:

1.) The question of transgender people is a question of how you treat a vulnerable group of people, whose lives are obsessed over in media and popular culture

Transgender people are more likely to be poor, more likely to be physically assaulted and more likely to attempt suicide. Despite attempts to create a panic about isolated cases of violence by trans people, trans people are vastly more likely to face violence than to do violence, in light of that, attempts to make moral panics about trans people are obscene.

The question of ethics is the question of how you treat the vulnerable people who won't retaliate or pay back the good or evil you do them. This is the most fundamental question, always.

2.) The question of transgender people and their lives is not primarily a question of metaphysics or semantics

Whether you think transgender people are “really” of their preferred gender matters little. My personal view is that words have multiple meanings, and there is no magic to their meaning, meaning is use. There is nothing semantically illegitimate about choosing to mean by the word “woman” that such and such person sees themselves in a particular way. It's just another way of using the word.

However, if you disagree with the above, and intelligent people can disagree on metaphysics, your obligations to respect transgender people do not change. This is because questions of ethics can't be decided with debates about metaphysics or semantics.

3.) The question of transgender people is not a question of whose supporters you find most annoying

Certainly some supporters of transgender people are annoying, have you seen the pic crew avi people on Twitter? Certainly also many opponents of transgender people are annoying “how many genders are there, why identify attack helicopter apache, why are generation Z confused about which bathroom use, haha v. funny”.

None of this should mean anything about transgender people. Transgender people are made of flesh and blood, not internet memes or buzzfeed headlines.

4.) The question of transgender people is not a question of whether you think all transgender people are “real”

It is sometimes suggested that a certain portion of trans people might be faking it for attention. Most often this is suggested of people who make relatively minimal changes to appearance and lifestyle and identify as non-binary.

I have no idea whether this is true on any significant scale, though human nature being what it is, I’m sure at least one person has tried it at some point. I really don’t care. There are far more harmful ways to seek attention. I would rather respect someone who doesn’t really need it than have disrespected someone who needed it desperately. I would rather assume authenticity and be wrong, then assume inauthenticity and wound someone.

5.) The question of transgender people is not a question of the psychological origins of transgender people

Personally, I don’t find Blanchard’s hypotheses on the origin of transgender people very convincing. But if he’s right? If some or all transgender people are autogynephilic and

on some deep level, unknown even to themselves this is what drives them to be transgender? That doesn't matter. Let people live their lives.

6.) Right now, the most pressing transgender question is how did we get to the point where the BBC is interviewing this Lily Cade as a respected voice on trans issues, and how can we get away from this precipice.

Miscellaneous thoughts on the right wing of politics

Conservatives and children

1- I cannot prove this, but I strongly suspect conservatives view having children as a way to cheat death, much more so than liberals do. They want to mold their children into copies of what they see as exemplary in themselves, in order to "continue their line". Certainly, every lunatic I have encountered online who talks unironically of "propagating their lineage" or some such is rightwing, and there's a surprisingly large number of them. In this theory, for the rightwing, the idea of their children being fully separate persons to themselves, free and autonomous to choose from an unlimited number of lives, is anathema, because it prevents the transmission of self in a way that avoids death.

2- Even though this desire to propagate self through the family is far from selfless, nonetheless connection with family gets coded as altruistic in our society. Family motives can be particularly dark in political practice, as they combine an occluding combination of selfishness and perceived altruism. Here's a comical example- Alex Jones, host of Infowars, explaining, in a way that presumably his audience finds acceptable- that he would butcher his neighbors in the event of an apocalypse to keep his daughters alive a little longer. It's very funny and presumably intended to be scandalous, but I tend to think it's only three or four steps from the conservative mainstream: after all, ten million people visit Infowars a month. Alex wasn't worried about his viewers finding this offensive, and that says something.

3- From a selfish perspective, the conservative wants to erase other modes of life so they can better mold their family. Thus the voice of selfishness says "do it". However, there is no counter-voice to hold them back with guilt, because they perceive themselves as acting out of altruistic motives- protecting their family. Both the angel and the devil on the shoulder have been coopted, and so there is nothing to restrain them. "In order to

protect my extended self-interest- family, I must eliminate. In order to follow the altruistic path I must protect my family and thus eliminate”.

4- It's often a conservative trope, but it is true to say that the most dangerous man when it comes to politics is the one who is absolutely convinced he's doing the right thing.

5- Remember, this was never really about saving the children, this is about saving my children and grandchildren, and thus, by extension, myself. Other people's children might be included to a limited degree if I perceive them as kin. That's why conservatives couldn't give a fuck about arguments like “saving the children means eliminating child poverty”. This isn't about a philanthropic principle, it's the expansion of the self.

6- I have nothing new to say here but: It doesn't take a genius to guess that as cultural diversity becomes ever more present through the internet, the conservative desire to shape their children's microclimate will become increasingly desperate and paranoid. This is one of the causes of the “all my children's teachers are groomers” panic. The kids are getting alternative attitudes and modes of life off the internet and off each other of course, but the illusion of control is an attractive idea. Left to their devices, they would absolutely demand that society be remade so as to remove alternative models and ways of being for their children. Their children will always find these models. There's a crack in everything, that's where the light gets in.

7- I read somewhere once, though I can't find the source, a pithy quote. The reason both the right and the left are absolutely convinced that they're losing is that they want different things. The left wants to control policy, whereas the right wants to stop the process of cohort replacement. They want to prevent the cultural alienation of each generation from the generation that succeeds it.

8- Being a sane adult in modernity means accepting that your children will be aliens to you. We know this because every generation has been alien to their parents, even the

Talmud says that man's great love is for his son, but his son's great love is not for his father but his own son. Alienation between generations is the way of things. Modernity can only amplify that. Technology can only amplify that. It's never, ever, going to change until there are no more humans left and we've all died or ascended.

9- It is not a coincidence that a section of the right is obsessed with protecting children from so-called (not actually) "groomers" at the exact same time as another, partially overlapping, segment of the right is obsessed with seducing and/or marrying children (the pedophilic section of the anime community, etc.). Both groups are interested in children for the same reasons viz, they view children as plastic that can very easily be molded into any shape they desire. Those that want to keep queer content away from children are afraid they will be molded by it. Those who are sexually obsessed with children or child-like characters also view them as malleable, and thus still "saveable". Among non-pedophiles on the right (and to be clear I am not claiming that right-wingers are unusually likely to be pedophiles), we see an echo of this desire, among those who specifically look for women who haven't been to college and haven't got much life experience. Also in the creepy view that women "hit the wall" at 20.

The right online: Yuck conservatism and the yuckification of politics

10- The online right has made as one of its main catchphrases: "I will not eat the bugs, I will not get in the pod". Many have pointed out that this phrase reduces politics to consumption choices, which is true, but notice also that it makes politics about disgust. About a moment of yuck. I have no particular interest in living in a pod or eating bugs either, yet I'd like to think that my politics is more meaningful than a yuck instinct. The slogan is just disgust, not love, not yearning, not even rage (although this is of course present in a subsumed form). Just disgust at consumption choices, like a toddler with food anxiety.

11- For what it's worth, psychological evidence that conservatives have always been more prone to feelings of disgust than the left is not hard to find. For example, it's implicit in Haidt's moral foundation's theory.

12- But we have seen the creation of "yuck" conservatism, a conservatism unmediated by any kind of supernatural or secular ideology (or at least any ideology very explicitly enters as a secondary moment). It's a conservatism just based on reactions like "What if my wife slept with someone else, yuck", "what if my kid was gay, yuck", "what if my daughter was a thot, yuck", "what if I didn't feel fully masculine, yuck". There's no theorization here of a celestial father that disapproves of these things. There's not even an attempt to tell a just-so story about how these things will destroy society. it's just a direct emotional reaction. The role of "yuck" seems to be more immediate than in previous conservative ideologies, and the role of a cohesive religious or socio-political worldview seems diminished.

13- Granted, there is a section of the right which is solidifying its connection with Christianity, but in the main, it's dropping away. The Christianity that remains increasingly becomes not a bedrock of the conservative belief system, but a mere synecdoche meaning, "the things I like". This is old ground, but what's semi-novel is that we're seeing a style of conservatism that rejects homosexuality, casual sex, etc. etc. without any kind of metaphysical foundation. A simple, unmediated sneer of hateful disgust exhausts the whole content. There's no particular reason these conservatives dislike, say, non-monogamy, they just find it ugly. It's the aestheticization not just of politics, but of ethics, and even of prudence (the sense of what will make one's own life go well).

14- If Walter Benjamin is right that fascism is the aestheticization of politics -the stripping away of ethical and prudential concerns and their replacement of an obsession with a kind of beauty- then the emergence of yuck conservatism may be a very bad sign.

An aestheticized, rather than theologized or even sociologized, ethics may be an ideological movement towards fascism.

15- Ultimately, this kind of “yuck, yuck, yuck” is fragile. It is fragile in the same paradoxical way masculinity is. Masculinity is meant to be hard and strong, yet it is treated as so fragile and vulnerable, that crossdressing once threatens masculinity. If your whole life is characterized by layers and layers of fears about things being yucky, you’re not an ubermensch, you’re a toddler with food anxiety. Such a personification of fragility! Isn’t there, ironically, something yucky about all this sneering and calling things yuck?

“Advertising conservatism”, “Waifu conservatism” and the inability to face reality due to being caught in a disgust trap where everything is sickening and scary

16- One of the most interesting manifestations of the aestheticization of politics on the right is fictionalism- the inability to cope with the world as it is, and the choice to inhabit fictional worlds

17- Although I do not fully understand it, fictionalism is not exclusive to the right. Liberals have their own version where they run to Harry Potter and the Marvel Cinematic universe. The unironic version of “Ruthkanda forever”. The liberal versions tend to be driven by emotions other than disgust though- most notably a kind of upset fearfulness at the cruelty of the world.

18- Among the very online right, fictionalism appears to be driven by disgust. Consider, for example, the waifu phenomenon in which internet users, disproportionately rightwing, will fixate on fictional women from Japanese cartoons. There’s no mystery as to why, they’ll tell you themselves. They like fictional women because they don’t have the flaws of real women, whether physical or “historical”. A selling point is that many

waifus don't have a sexual history, and these men are terrified of women who have slept with anyone before them.

19- A friend of mine put the dynamic this way. These people like sexy pictures of women. Unfortunately, they are so sickened by women who have had sex with anyone previously that they are disgusted by a real photo of a woman looking sexy- because the fact that she has taken a sexy photo is evidence that she has had sex previously. Fictional women are the "solution".

20-Of course the waifu phenomenon is also driven by a fear of rejection by real women, but even this tends to blend into disgust in interesting ways. See for example this random tweet that put it very directly:

21- But it goes far beyond the waifu phenomenon. Find some unironic conservative memes that say "retvrn" or "never forget what they took from you". Notice that they're rarely actual honest photographs of the past. They're usually ads from the past- idealized pictures of happy families made in the 50's, 60's 70's and 80's to sell products. These people aren't pining for a world that was, they're pining for a world that was advertised. Why? Because the actual world with its suffering, screaming, teeming, shitting, living, and dying masses disgusts them. This isn't really a revolt against the present. This is a revolt against reality from people whose sense of disgust has become so hypersensitized that the real world makes them nauseous.

Occultism and the aestheticization of belief

22- A lot of the very online right are getting into the occult these days. That takes a lot of forms. Vague references to the spirit world, and to demons, in a way that has become unhinged from Christianity. Belief in the power of curative potions and poultices. Strange talk of "manifesting" and "energy".

23- Occultism is a kind of perfect post-belief belief system for an aestheticized view of the world, because contemporary western occultism, post theosophy, and chaos magic, are basically the substitution of aesthetics for belief, of vibes. “What I want makes the world” occultism even shares the narcissism implicit in aestheticization, “what will be depends upon my taste!”. The westernized idea of the Tulpa, something that exists just because you want it to, belief preceding the world rather than the world preceding belief, is the ultimate summation of this view of things. Knowing is replaced with willing. Truth is replaced with appeal.

24- Unsurprisingly then, the very online right has embraced occultism with enthusiasm. It’s a boneless form of occultism, with reference to “forces” and “spirits” but few names or set stories. But then they managed to create a whole theory of politics without once talking about the mode of production, so why should we be surprised that they managed to make a whole practice of magic without ever actually theorizing the sinews of another world. Above all, occultism allows the right to externalize their sense of bad vibes and disgust by attributing it to “spirits”. Watch Cernovich here seamlessly move from “this makes me uncomfortable and sickened” to “this involves dark spirits”.

[In the original there were screenshots of tweets in which Mike Cernovich, a conservative commentator, attributes his bad feelings about transhumanism, metal music and horror movies to the dangers of evil spirits and the spirit world.]

25-I don’t think it can be emphasized enough that these people are fearful failures
deadly dull.

Back to the mainstream right: Post-neoliberal politics

26- Thus far I’ve been talking about the lunatic online right and their desire to kill reality and replace it with something that doesn’t make them feel disgusted. Now we’ll talk about something much saner. The “official” “sanctioned” right and their aim to kill the planet.

27- In post-neoliberal politics, everyone agrees on no serious efforts at redistribution. Even more sacred, everyone in the political class agrees on privatizing everything. The last remaining political vista, the last remaining thing over which we fight in the economic sphere, is whether or not people should have to pay for their externalities. Liberalism says yes, and conservatives say no.

28- In an environment where redistribution is off the table “not paying for your externalities” can take on a strange left-wing sheen as Pigouvian taxes -taxes on negative externalities- are basically just flat taxes that will not be compensated. This explains much of the “populist” flavor of the right on economics in recent years. In a situation in which the public had confidence in the power of the government to redistribute the public might have much more faith in measures like carbon taxes.

Hot and cold

29- The internet right can be seen as made up of two groups of active agents, the nihilists, and the opportunists. The nihilists and opportunists are opposites in an interesting way. To be clear, there are those on the right who don't fall into either group, but the ones in these two groups tend to be the doers and leaders, the others just sort of follow along.

30- The nihilists don't give a fuck about their own life. The real world disgusts them so much that they'd be happy to die in a flame of narratively-rich glory, transcending this world for the legends they much prefer.

31- The opportunists look at the world, look at their (real and imagined) skills and capacities, and look at the boomers and think “hey, why ain't I rich like the boomers”. So they look for the cheat code. Crypto looked like it might be the cheat code until it wasn't. They have zero interest in overthrowing the power elite en masse, rather they fantasize about being recognized for their unique talent and genius by a portion of the power elite, and invited to join.

32- The opportunists would never join the left because what they want is to have a prosperous family (and perhaps also plenty of chances to play the field on the side). They know that any kind of project to negate the existing power structure of society is not their best bet to secure for themselves a prosperous family. They're not wrong about any of this, from an individual perspective, trying to make the world a better place by taking the fight to almost all of the power elite simultaneously is not going to make your life go better for you. From this point of view, gambling your money on crypto or praying that Elon Musk notices your tweets is an infinitely better strategy for trying to change your life.

33- The opportunists aligning with the right is dangerous in a way that the nihilists aligning with the right isn't. The opportunists may seem non-serious, blinded by cupidity, but actually, they are serious in a way that the nihilists aren't.

34- The natural weakness of both these poles of the right is that they're shithouse at organizing. The opportunists can't organize because of their deep individualism. The nihilists can't organize because they're too unstable. Unfortunately, that doesn't slow them down all that much because the online right has something the far left doesn't have- a media infrastructure that organizes for them, that tells them what lines to spout, what strategies to adopt, etc. Both the right and the liberals leave us in the dust in this regard.

35- The left stands, in an odd way, in the middle between these two groups. They hate bourgeoisie society too much to want to join it. They're not filled with the same kind of all-consuming anger at life that makes the nihilists want to lash out at everything. Life and even bourgeoisie freedom, are beautiful things, but they are being parasitized/colonized by capital. Implicitly or explicitly they hold their own life lightly enough to risk it, for anyone who calls for a revolution- a metaphorical or literal revolution- is at least calling for an opportunity to wager their life.

Why hypocrisy is built into the right on questions of criminal justice

36- The right is perpetually getting caught doing this or that hypocritical thing about criminal justice. Endorsing harsh punishments for others, but not when their own friends or kin, metaphorical or literal kin, get caught. The reason is very basic and easy to understand, but I think it is still worth spelling out.

37- There is a sense on which this is not hypocrisy on behalf of the right. What the right believe is that there is a dangerous rabble who are not quite people. We must be cruel to this rabble. Exuberantly so. We have to draw hard lines, and the sharpness of these lines against some people who are relatively innocent is no argument against these lines- on the contrary- it is evidence that these lines are working. To use a cliché, the cruelty, or rather the excessiveness is the point. The mob itself is excess, only excess will control it. The mob isn't really human anyway. At worst the odd life imprisonment for someone stealing a biscuit on a three-strikes law is regrettable at best it is a positive good- culling the herd.

38- But because the right want these lines to be cruel and sharp against the ill-defined mob, it is no surprise that they don't want them used against people they actually consider people. Obviously, my kin and friends aren't part of the rabble, they are real people. The whole point was to be excessive against the mob, but tautologically you don't want to behave excessively towards actual people. Far from betraying its own principles by endorsing different standards of punishment, supporting different standards of punishment is the only way the right can faithfully express its view that only some people are people, only it can't quite say these things openly, so it comes across as inconsistent.

The correct line on cancel culture- revealed!

39- Cancel culture is real and horrible, it's even killed a bunch of people through suicide. There are real, vicious idiots who will think nothing about destroying lives over sincere mistakes and honest disagreement. The right do it too whenever they get the chance don't you worry.

40- My sense is that the left have, for the most part, targeted celebrities, microcelebrities and people with ambitions of becoming celebrities and microcelebrities. The right are not so discriminating, recently they have targeted a lot of individual teachers for example. These aren't hard rules though, everyone will go after anyone, to a degree it's the war of all against all out there. There ain't no one respecting the laws of Lashon Hara, or very few.

41- But the main thing to remember about cancel culture, right or left, is on the scale of politics, it is a relatively minor problem. I don't know exactly what the right policy is on COVID, but it's killed a million people so far. There are two million people in jail or prison in the US at any given time, and the lifetime chances of going to either at some point are much higher. Nuclear war, climate change, these are the real stakes of politics. Anyone who uses cancel culture as a reason not to be on the left is an idiot. Almost all the danger cancel culture poses is that it weakens the left, at a time the left desperately needs to be strong. I say this as someone who is as at risk of being canceled as anyone. As someone whose OCD makes them especially fearful of being canceled. This stuff could very well kill me one day, if anything drove me to suicide it would probably be this. Yet on the scale of the world's problems, it's as dust. That's the truth about cancel culture. It kills people, yet it is not a first-order concern. Again, politics is about wars and rumors of wars, starvation, imprisonment, mass disenfranchisement, homelessness, murder, and social murder in all its various forms. This is small fry.

42- Side point: I saw a great point about identity politics, and by extension cancel culture, on Twitter recently that I think bears repeating. A lot of the complaints about

cancel culture now are coming from the centrist establishment. The establishment has been blaming cancel culture on the left. but if you actually look at who has inflicted “cancellation”, it’s often awful centrists. The centrists created this HR-bullshit-identity-politics-complex, and now they’re blaming the left for it. Remember when Clinton was like “we could break up the banks tomorrow, but would that end racism”? A lot of this shit came from the center, and now the Matthew Yglesiases of the world are wringing their hands and complaining that we the left have chased the voters off with this stuff. Look at your own hands, sir! This is not to deny that there has been a real cancel culture, at times, in the left, but don’t let the right or the center try to worm out of their share of responsibility.

The big problem with the left

The title is a lie, I'm not going to tell you the big problem with the left- not exactly, or at least not in a particularly useful way. Wishing people would change en masse is a mostly futile gesture. Thus identifying a common personal characteristic of the big problem of a group is misleading. It's a bit like saying "if the left were able to fly faster than a speeding bullet we would win very easily"- it might be true, but if wishes were horses...

Nonetheless, the problem I'm going to identify here is a personal characteristic and I am going to make a useless wish that we would all change en masse not to have that characteristic. Whether the fault is just within us or within our stars seems academic.

Ado aside, the problem with the left is that we're not strategic actors. Instead, we act out roles. Let me give an example:

Person A on Twitter posts: "The coolest thing about the left is there's almost no legitimate way to learn or come around. If you weren't reading Marx in high school then you're a piece of shit liberal till death so why even try."

Person B on Twitter quote tweets person A, with this complaint: "Some people's politics are built around whether or not people are nice enough to them."

In other words, someone has put forward a complaint about the left that it is not welcoming enough. the quote tweeter responds by saying that, given the stakes left politics is addressing, demanding people be nicer to you is petulant. Person B is acting out the role of someone who is taking this Very Seriously.

I sympathize with the quote tweeter and I see where he's coming from- "people's lives are at stake, suck it up sweetheart" it's something I've often thought myself. The problem is that he's failed to apply his logic to himself. The natural continuation of what

he is saying is that if the stakes are so high, and if there are so many people out there who won't do the right thing unless you're nice to them, then you better suck it up, put a smile on your face and be nice to them rather than following the "very serious situation script" and getting mad at them for not already being on the right side of history.

I think that the thing people don't understand about radical politics is that it is still politics. Politics is the continuation of war by other means, especially the bit before the war starts in earnest where you try to gather together as many friends as possible. Do you know how politicians go around shaking babies and kissing hands kissing babies and shaking hands? There's no need to do that exactly, but we should be thinking along those lines.

To put it differently, the thinking in the above tweet is ought thinking- other people ought to be doing this rather than means-end thinking how to get other people to do this. There's no thinking through the logical interrelationships between the psychology of others, your goals, their goals, and the arrangements of things in the world.

A lot of discourse reminds me of a frustrating conversation I had with an old boss once. My boss complained that my coworkers were constantly getting a particular process wrong, he asked me what could be done about it. I suggested he remove the process or alter the process in some way so that particular mistake became impossible. His response? "No, they ought not to be making the mistake, how can we stop them from making the mistake". He cared more about the mistake not being his fault than he cared about preventing the mistake. It's very, very like a leftist who cares more about the fact that people should already be on the left, than working out a strategy to move them there.

If you're on the left, I urge think more like a chess player, like someone who has resources, has a goal, and deploys their resources to achieve that goal. Now it's true that

going too far down this route, and not thinking about other people in terms of their agency and moral obligations at all, can fuck you up. I grant that.

But it's a spectrum, and I'm 99% sure that, at present, this is the direction we need to move along it. Most important act on your objectives, not on the basis of a prewritten mental script of what you are supposed to do or feel. Be a goal directed agent in the world.

But then isn't this essay just another instance of the very thing I'm complaining about? Maybe, not exactly, kinda, I dunno, aww shucks.

A very brief argument for materialism about power

Materialism is roughly a view which places special emphasis on economic conditions—production, exchange, and consumption in the analysis of society. Materialism about power then is the view that social power has an absolutely central foundation in economic conditions. Through most of recorded human history, materialism about power would have been utterly humbug, a triviality not worth mentioning. Recently though, a strand of the left has developed which sees wealth, income, and class as just another facet of power. In theory, it argues that class is no more fundamental than say race or gender. In practice, it often goes so far as to treat class as less important than race and gender. This view is sometimes called intersectionality but that's a misuse of Kimberlé Crenshaw's terminology in my opinion(1). A better name would probably be kyriarchy theory.

Here's the quickest way to see that kyriarchy theory is wrong.

If you gathered the most powerful people in the world together, a disproportionate number would be white, a disproportionate number would be men, a disproportionate number would be straight, etc. But there's only one thing 100% of them would be—wealthy(2). Google Davos or G20 photos if you don't believe me.

Google some Davos and G20 photos if you don't believe me.

Footnotes

Footnote 1: "Intersectionality" is the observation that when a person is oppressed in two ways, e.g., through race and gender, their experience of oppression cannot be reduced to a race and a gender component, but rather these interact in complex ways. This is

completely true, and it's a real shame that the term is often stripped of its specific meaning and made into a synecdoche for identity politics as a whole.

Footnote 2: For those readers who are Marxists: A better way to understand the relationship between sex, race, and class is that sex and race help structure and organize class, adding differentiations within classes to create a working-class which is both politically divided and stratified in terms of its economic role. C.f. Selma James' *Classic Sex, Race & Class*.

Economic justice and climate justice are not metaphors: A response to Justice Creep by Scott Alexander

I. Cause I'm a Justice Creep, I'm a Justice Weirdo.

Scott Alexander has a new blog post up in which he complains about a phenomenon he calls justice creep. A move in language away from talking about how it would be nice to help certain groups or causes, to talking about securing justice for those causes:

Helping the poor becomes economic justice. If they're minorities, then it's racial justice, itself a subspecies of social justice. Saving the environment becomes environmental justice, except when it's about climate change in which case it's climate justice. Caring about young people is actually about fighting for intergenerational justice. The very laws of space and time are subject to spatial justice and temporal justice.

But I disagree. I don't object to Justice Creep. Regardless of whether it is useful -and I hope it is- I think that honesty compels a clear-eyed person to talk about many of these things in terms of justice, even in the narrowest conception of justice.

The mistake in Scott's article is assuming that these forms of justice are merely metaphors or analogies on criminal justice. Many of these are about justice in exactly the same sense that crimes are about justice- no metaphor required. Of course, they are also about being just in other senses- justice was never just about crime. For example, one can detect demands for social justice in the bible that go far beyond "wouldn't it be nice to help people", but nonetheless aren't framed in terms of the criminal law.

Nevertheless, yes, climate justice and economic justice- for example- are also about being just in the same way laws against murder are- no stretching of meaning is required. Unfortunately, this point often becomes lost due to something I call the legal veil. The legal veil prevents us from fully grasping the moral dimensions of actions that

are officially sanctioned. Instead, we assess them in a dreamlike manner- they become “regrettable”, “scandals” etc. but stop being crimes.

Let us briefly define terms.

An issue of justice, in the same sense of justice as criminal justice, arises when one person foreseeably harms another, or when one person threatens another with harm.

An act is unjust, in the same sense criminal acts are unjust, when the harm, or threat of harm, cannot successfully be defended as right or necessary e.g. an unjust act is an issue of justice per the above definition, that cannot be successfully defended.

II. The Iraq War under the legal veil

My point is perhaps best first illustrated by a point not from Scott's essay, but from one of the comments on the Reddit thread about the article. The author of the comment lists instances of "hyperbole" that they think are plaguing contemporary discourse. One of the examples is referring to the Iraq war as a "war crime".

I want you to imagine a post-apocalyptic society, say the burnt-out ruins of Sydney. There's a gang in Darlinghurst with 30,000 members and a gang in Haberfield with 4000 members. The gang with 30,000 members has knives and baseball bats and even a few score guns, the gang with 4000 members has big sticks. The gang with 4000 members is ruled by a brutish fellow- not well-liked. A rumor starts about him that he is building some guns- soon the gang with 4000 members might have guns of their own. The leader of the gang with 30,000 members drums up panic about this fellow's guns of mass destruction. Eventually, he leads an attack on the gang. A hundred members of the smaller gang die. It turns out the smaller gang didn't have guns.

There are, as far as I can tell, zero moral differences between the situation I have described and the Iraq war. To the degree that there is a difference, the Iraq war is

actually worse, because people had more opportunity to be civilized given that we're not in a post-apocalyptic hellscape.

So why do people think of calling the Iraq war a war crime "hyperbole"? Because our brains treat formally illegal things as beyond the pale but treat violations of morality under color of law as "issues" "scandals" "tragedies". We place a veil of law over them, which subdues the moral wrongs. But surely, on any consistent moral view, this veil of law is a fiction. I wrote about this in an essay once called A Katana, an iron bar and prison. The gist of it is, suppose you met a judge at a cocktail party who had, in your opinion, punished someone harshly in an obviously unjust way. you should treat that judge like you would treat a person who had beaten someone viciously with an iron bar for no good reason or more accurately, locked someone in his own basement for a decade. However, because the act was done through the "appropriate channels", they are protected by a moral blindspot. Moral maturity means recognizing that the moral veil is fiction, although perhaps -tragically- it might be necessary and useful, even though it is a fiction.

III. Macaroona

Now, to Scott's examples. Let's start with Climate Justice

There's a room. In that room are people sitting at different elevations. There are delicious macaroons laid out on the tables. Every time a person eats a Macaroon a little bit of a poisonous gas heavier than oxygen enters the room. Nonetheless, many people, disproportionately those at higher elevations, continue to eat Macaroons. The people at lower elevations are begging, pleading, screaming, and sobbing for the macaroon eaters to stop eating so many macaroons but it's just so difficult to coordinate everyone to stick to a macaroon budget, and besides, some people- including many suspiciously funded by the macaroon lobby- are arguing that the poisonous gas doesn't exist and... so on- I'm sure you get the metaphor. Talk of macaroon justice is talk of justice in exactly the same way criminal justice is talk of justice. When it comes to climate justice, just as

in macaroon justice, assault, property damage even murderous wrongdoing is afoot. However the veil of law - of official permission for the actions of governments, fossil fuel companies, and big polluters- stops us from seeing that.

This is why Scott's question about the little ice age is a bit silly. The little ice age wasn't unjust, just like spontaneous perfusion into the macaroon room of poisonous gas wouldn't be unjust. What is unjust is power players using state and corporate power to allow and even violently defend macaroon consumption, even after the stakes become obvious.

But does that mean neglecting those suffering due to natural variation in climate is not unjust, as in Scott's example of Mali? No, as we'll see now in our discussion of economic justice.

IV. Property rights are not leaving things as they are, they are an active imposition on the world

So what of economic justice? Step back and think about what property is. Stop thinking about it as an abstraction, and think about its real legal and social existence. Property is the right to exclude everyone else from using something under threat of violence unless they have your permission, or unless you transfer it to them as a gift or sale.

Now suppose you're dying in the snow. You walk up to a house and knock on the door pleading for shelter. A man with a shotgun greets you and tells you to fuck off. You do and die alone in the cold.

The distribution of property (which again is a social relation entailing the right to threaten someone with violence for using something) is a result of contrivances and power operations. Taxes and transfers and subsidies. At the end of the game of musical chairs, some people, whether through ill-luck, incompetence or in some cases the

malice of others, are left with nothing. Surrounded like tantalus with things that could help them, but that they can never reach because men with guns from a large organized gang (the government) prevent them.

Now you may argue that the current distribution of property is fair (doubtful- I don't think it meets anyone's standards of fairness given how arbitrary many of the rules are). You may argue that the current distribution of property is necessary from a consequentialist perspective (again, doubtful. Iceland has one-third the relative poverty rate of the US- I'm sure the relative poverty rate could be cut at least in half). However even if it is both fair and necessary, the question of whether there is economic injustice is still a question of justice, even if the answer is everything is just. This is because it is a question of forceful coercion. Asking whether Bob is unjustly exploiting Ellen is little different in principle from asking whether Bob is assaulting Ellen- whether the answer is yes or no. Yet again the veil of law has blocked our sight. Although the consequences may not always be life and death, the question of property is not unlike being forced outside to die in the snow- or at least to suffer greatly.

Edit: I want to be super clear on something even to the point of repeating myself slightly. Nothing I said above commits me to the view that all property rights are bad. It only commits me to the view that property is a right to violently exclude. Sometimes that may be a necessity, to ensure economic prosperity, for privacy or even to avoid danger (as with property over a nuclear plant) etc.

Nor am I saying that property doesn't exist, I'm just saying that to morally and politically evaluate property, you first have to get some distance from it by abstracting away from it, and analyzing what it is in real social terms. That means recognizing that property existence is not fundamental, but is made up certain kinds of arrangements of objects and people. What are those arrangements? In real social terms I think property is a relation of power over a thing, secured by power over other people- the power to

exclude them from that thing. This power over others is often, but not always, mediated by state power.

V.Of saints

Now as for saints and Scott's comment that this model of the world has no room for saints, I'd say this model of the world plenty of room for them. Heroes who stand up to moral criminals far more powerful than they are and risk reputation and life in the process are saints. As one commentator on the Reddit thread put it beautifully:

I'd think it would allow wise judges, paragons, righteous heroes, evil-slaying paladins, and so on?

To this list, we can add Martyrs. But we can also add in ordinary saints- saints of charity and not justice- the kind of saints that Scott complains this model does away with. Being charitable, kind, and merciful is nobler against a background of greed and malice.

But sainthood is about more than giving large sums of money philanthropically anyway. It's always been about giving when it hurts and giving when giving takes courage. I'm reminded of the Widow's mite:

"He sat down opposite the treasury and observed how the crowd put money into the treasury. Many rich people put in large sums. A poor widow also came and put in two small coins worth a few cents. Calling his disciples to himself, he said to them, 'Amen, I say to you, this poor widow put in more than all the other contributors to the treasury. For they have all contributed from their surplus wealth, but she, from her poverty, has contributed all she had, her whole livelihood. Mark 12:41-44

This quote from Archbishop Hélder Câmara

“When I give food to the poor, they call me a saint. When I ask why they are poor, they call me a communist”

And strangely, this quote from Lowell, repeated extensively in *Unsong*:

But the soul is still oracular; amid the market's din,
List the ominous stern whisper from the Delphic cave within,—
'They enslave their children's children who make compromise with sin.'

But we've all made our own compromises with sin haven't we? So please don't take this as holier than thou preaching.

VI. Am I calling for blood? No, but others might.

Am I suggesting piercing the veil of law to a permanent end? Am I suggesting Nuremberg-type trials for, for example, "climate criminals"? Eh, probably not. But this isn't because there is no consistent standard under which these people deserve Nuremberg trials, it's just that I'm both merciful and practical. I don't like the idea of punishing anyone. I'm the kind of person who goes to bat for violent criminals, even the sort of violent criminals most of the left loves to hate. However, if I had a great deal of power, and if I were more vengeful than I actually am, it would be brutally apparent that there is zero mottos and baileys here. This is very directly about crime, and by clear implication, punishment.

Others that come after me and my generation might not be so merciful. Something “climate criminals” would do well to remember, especially those of them that are likely to be alive in thirty years or more when the worst effects have set in.

VII. Final thoughts on ownership in response to a reader.

Edit: This response is quite long because I want to be very precise. A reader writes in relation to my analogy of a man being forced into the snow:

You don't say what we're supposed to take from the analogy, but it seems like the implication is that this is unjust. But it would not be unjust by most normal understandings of the term. The man's actions are, under many moral frameworks, immoral, but that's not the same as being unjust.

But it seems to me that of course, this matter relates to justice. You're being forced out into the snow to die. Generally, we'd call this murder. Being murdered is an injustice.

Now I don't know why the reader thought this wasn't an instance of injustice, but I can guess. They've assumed the house belongs to the gun wielder, ergo the gun wielder has a right to decide whether to let you in or not. Their action might be immoral, inasmuch as it is a refusal to extend charity, but it is not injustice inasmuch as it is not a breach of your rights. They have the right to the house. Things get tricky here because our intuitions about private property are very deep. I'm going to try to let you look at them from a distance.

The first thing to note is that I never said the house belongs to the gun wielder. That was a deliberate omission. I wanted to strip away all concepts of property. It is very interesting that the Redditor mentally inserted this though.

Before we get to the questions of the morality of property, there are only objects in the world, and people claim those objects by the threat of violence restricting the liberty of others to use them. Ownership is a social convention to use violence on people who try to use stuff that isn't recognized, by the social convention, as their stuff. In the first instance -looking at the world without concepts of property-, this guy is a guy who is

driving people out into the snow to die. Hence, prima facie, his actions are murder, unless they can be justified in some way.

Issues of economic justice are just that on a grander scale- debates over whether the violence used in pursuit of “ownership” are defensible in some particular case, or are instead indefensible restrictions on people’s liberty to use what they want to.

Now of course you could believe that private property exists as a moral right on top of this world with people, objects, and guns. You could also believe that private property means the man with the gun hasn’t acted unjustly so long as he owns the property. Fair enough. But that doesn’t stop being driven out onto the snow from being an issue of justice. It just means that it’s an issue of justice that resolves as “no injustice was done here”. The owner is charged with injustice but gets off on a private property defense. The accusation of injustice is still not a metaphor.

On the other hand, people who don’t accept that anyone has the right to force people out into the snow [e.g me] will regard it as an issue of justice in which grave injustice has been done. I don’t think the man has any real moral right to the house in the sense needed to force someone outside, there’s just a social fiction that he does. That social fiction allows him to rationalize chasing someone into a place where they will die but makes no moral difference.

My personal view is that some forms of ownership -violent restrictions on the liberty of people to use things- are defensible, at least at this stage of history, but only if they ultimately serve a consequentialist greater good. Hence forcing someone outside into the snow to die just because you don’t want them to sleep in your house is murder, because it doesn’t serve the greater good. No laws change that, and in the true moral sense of ownership- a moral right to exclusion- you don’t own your house in that way.

Brief observation on the political economy of the welfare state

It is often observed by conservatives that the existence of a welfare state reduces the incentive to have a job. Surely, this must be true in at least some cases. Of course it may also provide people with a means to pull themselves out of poverty, take greater entrepreneurial risks etc. counterbalancing this incentive.

It occurred to me though that there's another incentive effect which goes in the opposite direction, much less frequently remarked. The existence of a welfare state creates a political incentive for the middle and upper class to support policies to reduce unemployment & poverty wages in order to reduce their tax bills. The existence of a welfare state makes your poverty my financial pain through taxes, incentivising political support for job creation, investment etc. by the upper and middle classes.

The nasty side of my leftism

I'm pretty leftwing. I would like to think that my politics are motivated by compassion, and sometimes they are, but also sometimes there is cruelty to my politics. I don't like this side of myself, but it is what is.

Let me explain with an example. In the US there is a long-running debate over the minimum wage. This debate is normally framed in terms of whether a higher minimum wage would be better for the poor. However there has always been a strain of conservatism- usually hidden in official pronouncements, but quite visible in Twitter threads etc.- that holds that the real problem with raising the minimum wage is that low skilled workers simply don't deserve to be paid that much- especially if it comes at the expense of those with higher skills, but even if it doesn't. People who hold this are often the same people that yell at and insult waiters.

My visceral response to this is partly disgust at how callous it is, but there's another feeling I have as well, contempt. To my mind what these people are doing is engaging in a status competition with minimum wage workers. They don't want minimum wage workers to get more, because they feel that reflects poorly, vis a vis their relative desert.

And the cruel, sneering part of me finds it absolutely contemptible that anyone would set as their competition in life minimum wage workers. What glory could there possibly be in aiming to cleanly exceed a minimum wage worker? What kind of slug would hinge their reputation on such a paltry form of financial success. That's your long term plan for glory? A kitchen renovation that they have not? Will they sing your praises in a hundred years because you had a beach house and your waiter didn't? Will epics be written about your SUV?

Again, I'm not proud of this. Fortunately, this contempt doesn't extend to the minimum wage workers themselves- since they might be reaching for glory in domains other than

the financial, and besides of which, they aren't being nasty to anyone. Still the little dictators trying to stamp down the floor make my blood sing scornful laughter.

Pharmakos

Wikipedia explains the tradition of the Pharmakos as follows:

“A slave, a cripple, or a criminal was chosen and expelled from the community at times of disaster (famine, invasion or plague) or at times of calendrical crisis. It was believed that this would bring about purification. On the first day of the Thargelia, a festival of Apollo at Athens, two men, the Pharmakoi, were led out as if to be sacrificed as an expiation.”

(Remember that it is very plausible that a large majority of those regarded as ugly or criminal were poor. Certainly beauty and the aristocracy were conflated in ancient Greece, and poverty and criminality have always been associated)

It also mentions a bunch of interpretations of the ritual:

“Walter Burkert and René Girard have written influential modern interpretations of the pharmakos rite. Burkert shows that humans were sacrificed or expelled after being fed well, and, according to some sources, their ashes were scattered to the ocean. This was a purification ritual, a form of societal catharsis. Girard likewise discusses the connection between catharsis, sacrifice, and purification. Some scholars have connected the practice of ostracism, in which a prominent politician was exiled from Athens after a vote using pottery pieces, with the pharmakos custom. However, the ostracism exile was only for a fixed time, as opposed to the finality of the pharmakos execution or expulsion.”

As the kids say, there's a lot to unpack here. A low member of the community becomes the scapegoat, then is excreted, but before that excretion is raised to a sacral state. A lot

has been written about the concepts of purification, divinity, ostracism, scapegoating etc. involved. The archaic parallel with the figure of Christ in the Christian mythology is obvious (moreover it is said that Christ will be ugly in the book of Isaiah and the early church held as much).

But what most discussions never emphasize to my satisfaction is that the pharmakos ritual is not merely an assertion about the boundaries of “the community” as an undifferentiated mass, but is an assertion of ongoing status and class hierarchies in the community. At least to a degree, what is being asserted is material and social hierarchies. The ritual communicates:

- *That slaves, criminals and the ugly are marginal and can be cast out at any time. The fortunate thus tell themselves the lie that they don't need the less fortunate, or at least, tell the less fortunate this lie.*
- *That a slave, criminal or ugly person who “gets above their station” by being honored (e.g. through feasting) is a paradox that cannot exist in the community long, but must then be cast out.*

Some have admired the ancient Greeks for their assertion of the right of the weak to trample the strong, but when I look at this ritual, I don't see the rule of the strong. I see weakness, and a paranoid fear that the lower classes will see that weakness.

I think there's a continuation of something like this pharmakos ritual today. People want the spectacle of “bad” (i.e. unsuccessful) people suffering, placed into a kind of internal exile (homelessness or economic precarity).

But that always makes me laugh because: The truly strong contend against the strong, not against the weak.

A bit more ancient Greece to bring the pharmakos metaphor out. My very first lesson at university was on ancient Greece. Our tutor sat us down and talked about everything from our University's motto to the Godfather in order to bring out some themes about the Homeric period. But the thing that always stuck with me was when he told the class:

People think that the strong take things from the weak. In reality, only weak pretenders to strength do this. The truly strong assert their strength by giving to those weaker than themselves, not taking.

I try, always, to think in terms of charity and love, but when the dark half of me rears itself, it also often wants to give charitably but for a very different reason. Maybe you can see why my vicious nature thinks even worse of those who want to bully and crush minimum wage workers than my virtuous side does.

Everything is negotiable on the right (and left)

I've heard people express the following astonishments:

“Why are conservatives arguing for everyone to go back to work? Aren't conservatives supposed to have stronger disgust and fear of disease instincts?”

“Why do conservatives so often viciously criticize Democrats who had military careers? Aren't they supposed to respect troops?”

“Why are so many conservative commentators criticising doctors and nurses online, don't they venerate first responders?”

“Why are conservatives tied to capitalism? Isn't it corrosive of established traditions, family etc.?”

What they're not understanding is that while individuals may care deeply about principles in politics, as a whole, politics is about coalitions not ideas. Ideas are wielded strategically on behalf of coalitions. The right represents a coalition of the powerful (and in our society, the core of that coalition is necessarily capital and capitalists). The left represents a coalition of the relatively powerless. Any principle claimed by either the right or the left will be dropped, at least by a majority, if it conflicts with the interests of their coalition.

This is why I'm on the left. On average, the demands of the powerless will be more righteous than the demands of the powerful. There are diminishing returns to status, money and power in terms of what it can do for you, thus those with relatively little status, money and power will benefit more from gaining some extra than the powerful will benefit from clinging to it, or grabbing more of it. If you want to get involved in politics with open eyes, the first step is to ask yourself not what ideas are more right but

which groups, in the abstract, are more likely to have just claims.

If you want any leverage on historical processes you've generally got to cling to one coalition or the other and try to advance it and steer it. Sometimes you've got to try and steer it quite sharply, but not so sharply you're flung off. As a rule, the world is governed by people, not by ideas.

A katana, an iron bar, and prison

1. The facts, the law and our brief

In the early hours of August the 10th, 2018, Jett McKee broke into the home of Hannah Quinn and Blake Davis. McKee pulled a gun (only later revealed to be fake) on Quinn & Davis and demanded money, he then knocked out Davis with a knuckleduster blow to the forehead.

McKee fled the home with Quinn in pursuit. Davis woke up, groggy, and chased after them both, carrying a katana. He then claims, not implausibly, that he found Quinn cornered by the apparently gun-wielding McKee, and so he struck McKee on the head with the katana, killing him. There is some dispute over exactly what happened in the confrontation outside, with the prosecution saying McKee was simply running away. The judge's conclusion was that McKee had not posed any immediate threat to Quinn when he was struck with the killing blow. Whatever the jury thought had happened, they did not believe it qualified as self-defense. Thus, under section 418 of the NSW crimes act, the jury presumably either thought his conduct was not properly:

“to defend himself or herself or another person”

Or it was not:

“...a reasonable response in the circumstances as he or she perceives them.”

The couple then panicked and went on the run before being eventually apprehended by police. Blake Davis was found guilty of manslaughter, but not of murder and sentenced to five years and three months imprisonment, with a non-parole period of two years,

nine months. Hannah Quinn was found guilty of accessory after the fact and was given a community corrections order.

We'll give the jury the benefit of the doubt and assume manslaughter really happened and was proven here. The sentence of the judge was, in many ways, not unreasonable given prevailing legal standards (anything less might have been reversed on appeal), and very welcome given the almost comically aggressive way the prosecution pursued the charge.

Nonetheless, I think this is a great place to step back and think about why we send people to prison, and how we can reduce that number.

I won't bore you by reviewing in any great detail the case that can be made that Mr Davis shouldn't have been imprisoned but here is the case in brief: Whatever a reasonable response might have been to the circumstances, he had just woken up after being knocked out. He was presumably scared and confused, both because of the nature of the situation, his own injuries, and his prior mental illness (PTSD). The character of his crime means he was very unlikely to repeat it. A severe non-custodial order would get the point across.

I want to change the way we look at prison. More generally, I'm concerned that we don't really see violence enacted by the state properly- it is veiled behind law and politics, and that makes us not properly understand it as what it is, violence just like any other violence. Like violence generally, it can be necessary, but before we can decide that it is necessary, we must do our due diligence by grasping it as violence. Once we've done that, it will become very clear that the use of prison needs to be restricted. What I'm going to propose is a sort of mental exercise, in the form of a thought experiment, intended to make us confront social reality without the veil of custom, depersonalization, abstraction, etc.

2. The iron bar

Put yourself in the shoes of the judge. You've just decreed there is to be a custodial sentence. However, the officer of the court nervously clears his throat.

"Your Honour, I have just been informed by the prison administration that there is no room left in the prisons. However, our government, in their wisdom, have devised a cost-saving alternative."

The officer pauses and pulls out a large crowbar from a bag at his feet.

"Being beaten by an iron bar severely enough to be hospitalized is roughly as painful as a multi-year stint in prison. Therefore, her majesty's government has determined that in cases like this, beating with an iron bar will be substituted for imprisonment. Medical specialists will be on hand to carefully supervise the beating, to ensure that, while traumatic, it will not be lethal. Since you have assigned the sentence, you shall carry it out. Unless, of course, your honor would like to revise your sentence? Perhaps a large fine or a community corrections order with stringent restrictions would be sufficient?"

What I am suggesting is that if you wouldn't be willing to beat a person severely with an iron bar, you shouldn't be willing to send them to prison.

We're saying, for the sake of the hypothetical, that an iron bar is just as frightening as imprisonment (it's an equal deterrent), and incapacitates the victim as much as imprisonment from future crimes (say, by causing permanent damage to the musculoskeletal system). We're also saying that the permanent damage and difficulties caused are, on average, the same as being sent to prison.

Certainly, a look at the literature on the effects of imprisonment suggests that this doesn't make imprisonment out to be worse than it is. If anything, it grossly underestimates the effects and risks of prison relative to serious physical assault.

Evidence from America suggests that in that country, one year of prison reduces life expectancy by two years, and that, depending on high-school diploma status and race, being incarcerated for the first time reduces lifetime earnings between 267,000 and 1 million dollars.

I'm not an expert on this literature, but I think both commonsense and survivor testimonies would tend to indicate that going to prison is extremely bad. If anything, it is quite a bit worse than getting thrashed with an iron bar.

I think that almost no one, save a few sadists, and perhaps some of the relatives and close friends of Mr. McKee would be willing to beat Mr. Davis with an iron bar. Now if Mr. Davis were, for example, serially violent, people might reluctantly agree to do it if the alternative was for him to simply walk free or face a comparatively mild community corrections order, but they would do so with "fear and trembling", conscious of the dreadful choice they were making, and of the impact, it would have on their own psyche.

My fundamental claim- which is debatable, but which I will not get into debating here- is that the foreseeable consequences of your action are what you are accountable for. Perceptions of increased or reduced "brutality", abstracted from the actual consequences of an action, are irrelevant to its morality. Above all, I'm saying that if being directly exposed to the consequences of your own actions, and having to cause those actions in a more direct sense, would make you reconsider those actions then you should reconsider them without delay. Don't send people to prison, or condone sending them to prison, if you wouldn't be willing to beat the shit out of them.

3. The legal veil

Above all, what I want to rip through is something I call the legal veil. Our different reactions to prison vs beating someone with an iron bar are just one instance of the legal veil. Let me introduce it with an example.

Suppose a politician admitted, on tape, that the main reason they were introducing some bit of legislation that would jail thousands of drug users was politics. Oh, there were some considerations of public health, to be sure, but the overriding reason was to appease the police union and shore up hardline tough on crime voters.

There certainly could be a big reaction. It might even be talked about internationally. Very, very few people, however, would treat it as morally identical to the discovery that the politician in question was a serial kidnapper and torturer even though they involve the same kinds of harms and the same kind of malice and aforethought. This disparity in treatment is what I call the legal veil. We don't see officially sanctioned actions as equivalent to private action.

The legal veil has a number of different components to it. One is that legal actions are often, or perceived to be, "cleaner" and less brutally direct than things that do similar harm- hence the difference in our reaction to putting someone in prison and beating them with a steel bar.

But there's another component- the perception of legitimacy or authorization. I'm not denying legitimacy is real, I just don't see why it should excuse actions that are clearly not bona fide striving for the common good. This aspect, of the perception of legitimacy, also relates to iron bars versus prison inasmuch as if our society regularly authorized beating with iron bars, you'd probably feel a lot less squeamish about it.

There are some legitimate reasons why we should cut politicians, judges, senior public servants, etc. some slack. For example, people in these positions make an extraordinary number of decisions involving life and death and are therefore bound to get some wrong. My intention then isn't to excoriate for bona fide mistakes- the legal veil probably leads us to the right conclusion then in some cases, even if for the wrong reasons.

But often those who benefit from the legal veil are actually more culpable than those who do similar things for other reasons. Consider again, for example, the politician who extended the war on drugs for political reasons. He, unlike the man who kidnaps and tortures for his own sadistic pleasures, has made specific oaths and promises not to do that sort of thing. Arguably he's worse then. Now granted, if you had to pick one of them to babysit your kids you'd probably go with the politician, but this is a judgment of dangerousness, not of moral depravity.

Think concretely, specifically, and honestly about the exercise of power.

I.

GDP undervalues government services

Gross domestic product can be calculated in many ways, but one of the most common is:

$(\text{Government expenditure}) + (\text{Consumption}) + (\text{Investment}) + ((\text{Exports}) - (\text{Imports}))$

To see why government expenditure is under-counted consider two countries, one of which provides all of its health care through the public sector, the other through the private sector.

In the latter case, health care is counted in GDP as the amount that consumers pay the private provider.

In the former case, health care is counted in GDP as the amount it costs the public provider to provide the service.

Assume it costs the same amount to provide the service in both countries. This is actually a generous assumption, governments can usually provide healthcare for cheaper than the private sector.

Assuming further that the profit margin of the private provider is not zero, it follows that:

$(\text{Price of the private health care services}) > (\text{Cost of the public healthcare services})$

In the simplest terms, government expenditure is devalourised as a contributor to GDP because it only includes the cost of provision, not the profit margin.

II.

The political implications are obvious. Insofar as GDP is a metric that voters (and investors) care about, politicians are incentivised to focus on private, rather than public growth.

III.

There is a long list of common complaints against GDP. It doesn't include damage to environmental assets, or resource depletion. It doesn't include unpaid labour- and therefore especially discounts the contribution of women. It pays no attention to distributional concerns, and treats an extra-dollar going to a billionaire as just as important as an extra dollar going to a pauper who truly needs it. Given these problems, many of which seem much more serious than a little under-counting of government activity, why make a big deal out of this?

The answer is that GDP is, in a sense, honest about these problems. Everyone knows these faults with GDP, many people can figure them out themselves after a minutes thought about the definition of GDP. The problems a measure wears on its sleeve are less grave for that measure than the problems a measure hides. The problem we have discussed here is less obvious than the others, because at first glance it appears that GDP does allow for government expenditure, yet the worm hides deeper in the apple.

IV.

There are three natural ways to deal with this problem that I am aware of. All have their pros and cons:

1. Value government services at the sales price of equivalent private services.
2. Value government services using willingness to pay analysis

-
-
3. Inflate the government expenditure term by a multiplier equal to the average profit margin.

Of these my favourite option is probably (3), although it sounds a little eccentric at first. Option (1) might unfairly overvalue government services if the private provision of certain goods (like roads) is very underdeveloped in relation to the government, and government provision benefits from economies of scale. Option (2) effectively gives the government a sort of monopoly power, since it's based on the absolute maximum people would be willing to pay, with the only other alternative being losing it altogether. It might be tempting to say "let's just not adjust for it at all, it's too hard to find a fully non arbitrary answer." But not adjusting it and accepting the biased status quo is equivalent to a deeply arbitrary and pretty steep downward adjustment.

I don't know how to tell you that politics is about murder

There have been numerous blow-ups during the Democratic primary about words and behaviour. The interminable debates about Joe Rogan. Elizabeth Warren partially attributing her non-endorsement of Sanders to snake emojis. Endless tone policing of Bernie surrogates like Briahna Joy Gray. The details have been stripped like desert bones and are, in any case, irrelevant.

I'm concerned about the enthusiasm with which these arguments have been taken up, and I think you should be concerned about this too because the substance of politics is murder. It's about the avoidable deaths—and the enslavement—of human beings. Any discussion related to politics that is not directly related to life and death, or human immiseration, should be of secondary importance. Almost 1% of the US population is incarcerated. 26,000 Americans die a year due to lack of health insurance. There aren't even good figures on the number killed by American sanctions each year.

Maybe this all sounds a bit repetitive, but sometimes you've got two options, silence and repeating yourself about a message that already should be heard, and it's better to repeat yourself.

This isn't an essay to tell anyone they must believe as I do. Maybe voting for Joe Biden is the right way to stop the ghastly procession. I do not accept as legitimate however the idea that we should decide that on the basis of this bullshit. If you are making your choices on the basis of snake emojis or pretending you are, you are doing the wrong thing and must stop. The questions of power cannot be evaded. Total political apathy would be more honest and clean.

It happens on the left too. For example, Mike Bloomberg's personal behaviour has often been appalling. Nonetheless, it bothers me a lot that we paid more attention to his personal cruelties than to the thousands of people who died avoidably, and the

hundreds of thousands who were harassed by the police unnecessarily, during his mayoralty. Why are we drawn back into these personalities all the time, it's not even like they're particularly interesting personalities.

Actions that unnecessarily kill people don't somehow take on a different moral meaning just because they're done by a politician *Thumps table* *Thumps table*. The inability to grasp that murder through politics is no different to hiring a hitman really colours our perception of the moral realities of politics.

People have a great deal of trouble giving murder and wrongful death the attention it deserves unless it happens in a spectacular, made for television way. Our natural priorities when it comes to politics seem to be:

1. Big events with relatively small death counts (e.g. 9/11)
2. Meaningless celebrity gossip type bullshit
3. Stuff that actually kills a lot of people (sanctions, lack of healthcare etc.)

This tendency has long worked against justice. Mark Twain remarked of the French Revolution:

“THERE were two “Reigns of Terror”, if we would but remember it and consider it; the one wrought murder in hot passion, the other in heartless cold blood; the one lasted mere months, the other had lasted a thousand years; the one inflicted death upon ten thousand persons, the other upon a hundred millions; but our shudders are all for the “horrors” of the minor Terror, the momentary Terror, so to speak; whereas, what is the horror of swift death by the axe, compared with lifelong death from hunger, cold, insult, cruelty, and heart-break? What is swift death by lightning compared with death by slow fire at the stake? A city cemetery could contain the coffins filled by that brief Terror

which we have all been so diligently taught to shiver at and mourn over; but all France could hardly contain the coffins filled by that older and real Terror — that unspeakably bitter and awful Terror which none of us has been taught to see in its vastness or pity as it deserves.”

My crackpot theory is that it comes from a kind of illusion where some people believe that there are only two kinds of people in the world—people they know, and celebrities (inclusive of politicians). This is because these are the only people we regularly encounter.

A less kind theory is that this gossipy rubbish is conscious or semiconscious dissimulation. People talk about this stuff because if they talked about the substance of politics when defending their choices, they'd have to admit they aren't half as left-wing as they pretend to be.

For communism and against foreclosure on the future

I. The problems of communism

By communism here we mean a system in which the principle of: From each according to their ability, to each according to their need is implemented as the sole principle of economic distribution and production.

There has never been an advanced communist society—only societies that aspire to communism in the long run. Money still existed in the Soviet Union, China and Cuba. The state paid some people more than others, and not on the basis of their extra need. It is not clear that there has ever existed an advanced society that even had the capacity to build communism.

There are I think two main objections to the possibility of a technologically advanced communist society, viz:

1. The incentive problem
2. The calculation problem

The incentive problem is the problem of making sure the work gets done, especially boring, dangerous and stressful work. If everyone gets paid on the basis of need rather than effort, why would anyone want to do these kinds of work?

The calculation problem is more technical, consult the Wikipedia article “The Economic Calculation Problem”.

These are legitimately tricky problems. One way to respond to these difficulties is to give up on communism—for years I did just that, in favour of social democracy or

democratic socialism. Another response comes, ironically, from a right-libertarian I knew in University.

II.A libertarian intervenes

I was putting to this libertarian the technical problems with anarcho-capitalism. These include the provision of public goods and management of externalities. I argued that these problems showed anarcho-capitalism was either impossible or undesirable. His response to this was twofold:

1. A) Anarcho-capitalism is a regulatory ideal—an organising concept for political action. Something can serve as such an ideal even if we don't yet know if it is possible.

2. B) The world we live in would have been unimaginable through most of history. To foreclose on the possibilities of the future would be foolish. Tremendous changes in technology and social institutions await.

Although I'm certainly no anarcho-capitalist, I think this is not a bad defence of aspiring to anarcho-capitalism, despite the technical difficulties.

These days, this is how I feel about communism. There are forces that push in the direction of communism. Communism is a good regulatory ideal for those forces. It unifies the most radical progressive tendencies in society. Will it ever happen? Predicting the direction of social and technical advances in advance is impossible so it would be foolish to rule it out. Inasmuch as it centres human needs in the productive and distributive process it is a good aspiration.

Note: Consider especially advances in AI and transhuman enhancement.

We need to be exact about what is being proposed here. I'm not arguing "No one can

prove communism is impossible, hence we should believe that it is possible.” That would be to substitute faith for reason. Instead, the idea is that “We don’t know whether communism will ever be possible but it may well be. Moreover there are reasons to think it is desirable, and we can take steps to try and approximate it better in the present, in this sense of aspiration it is right to aspire to it.”

The danger of an aspiration like communism is that it can overwhelm more specific social plans for the coming decades. So long we avoid this danger, we should happily regard ourselves as communists.

III. The line between revolution and reform

I wanted to talk about another issue which is quite logically distinct, but which I regard as emotionally linked: reform vs revolution. What I say here will not be new to anyone who has thought about this for a few years, but I remember that there was a time that I didn’t understand it, so maybe it will be helpful for some.

There is no absolute divide between revolutionary and reformist strategies. There are only degrees.

The most obvious reason this is true is that winning reforms can increase revolutionary morale and organisation. Conversely, the threat of revolution can win reforms.

The deeper, often missed point is that winning elections can create legitimacy for an anti-capitalist movement. If the deep state then responds through a coup, a revolution can happen in the defence of an elected government—revealing and hopefully defeating the true nature of the repressive apparatus(1). If successful, such a revolution can then dismantle the existing capitalist state. For this reason, revolution and electoral victory shouldn’t be counter-posed strategically, at least not in an absolute sense.

Engels describes how revolution can be a defensive option, once the state represses legal
Methods:

“[...] Be that as it may, for the time being it is not we who are being destroyed by legality. It is working so well for us that we would be mad to spurn it as long as the situation lasts. It remains to be seen whether it will be the bourgeois and their government who will be the first to turn their back on the law in order to crush us by violence. That is what we shall be waiting for. You shoot first, messieurs les bourgeois. No doubt they will be the first ones to fire. One fine day the German bourgeois and their government, tired of standing with their arms folded, witnessing the ever increasing advances of socialism, will resort to illegality and violence.”

The egalitarian past (and future?) of politics

Let's say that politics is the formation and use of coalitions for determining matters of social concern within a group.

I was listening to a talk about human evolution as a process of finding a cognitive niche by Andrew Whiten when he made two critically important points which, while I was aware of them abstractly, I'd never previously considered as points about the inherent character of politics.

1. Among early hunter gatherers, we have strong evidence to believe that political coalitions were far more likely to be used to oppose the ascension of a person to a position of social dominance than support it. This is well known- there's a tonne of research on counter-dominance strategies among hunter gatherers, and the point that most early hunter gatherer groups were politically egalitarian in addition to economic egalitarianism is well understood.

2. Even among chimpanzees, some of the primary uses of coalitions include supporting the interest of groups of weaker individuals against a single stronger individual, and acting as king-makers, preventing a single male exercising absolute dominance.

It occurred to me that these days we generally view politics as a tool of the powerful, and yes, it is. In the very beginning though, we have reason to believe that politics was an innovation of those individually less powerful- the formation of coalitions to control charismatic or physically powerful individuals.

In other words, pervasive use of politics, was once a feature that separated us from certain other types of animals, making us far more egalitarian than, for example, chimps. Almost unimaginably, the very feature of our behaviour which made early

societies so much more egalitarian than many other primates would one day enable certain individuals to possess billions of times more wealth than others. Our challenge then is not just to explain why human societies have moved from the egalitarian to the inegalitarian (and in some cases, part of the way back again), but to explain why and how the primary use of politics was transformed from counter-power to power. How did the orientation of politics become inverted from its original use? Reevaluating our conception of the origins of politics might enable us, as people concerned with a more egalitarian world, to relate to politics differently, and with less revulsion, viewing it most naturally as 'our' tool not theirs.

The alternative conception of politics has, after all, never quite gone away. Even today, leaders of human groups are required to at least pretend to rule in the common interest, a throwback inherent in the way political power is structured, as an exercise in the perceived interests of the many who form a coalition.

Trying to envisage a road to equity is a fearful task, but we shouldn't be too afraid of that maligned tool, politics- ancient friend of the many and foe of the few.

Mistaken Identity and misunderstood interests: Haider and identity politics

I just finished “Mistaken identity” by Asad Haider, and like anyone who has just finished a good book I’m a proselytiser for it. My aim here is to draw out one thread of its multifaceted arguments, that the whole of the working class share a joint interest in abolishing racism in a way that is not recognised by what is often called identity politics.

Like Haider we will only be discussing racial identity politics here, and focusing particularly on the problem of white supremacy in America.

Consider this quote from Ignatiev reproduced in *Mistaken Identity*:

“To suggest that the acceptance of white-skin privilege is in the interests of white workers is equivalent to suggesting that swallowing the worm with the hook in it is in the interests of the fish. To argue that repudiating these privileges is a “sacrifice” is to argue that the fish is making a sacrifice when it leaps from the water, flips its tail, shakes its head furiously in every direction and throws the barbed offering.”

What Ignatiev is suggesting—and Haider concurs—is that rather than accepting that the white worker benefits in any unequivocal sense from racism we should consider a more complex view. As Haider puts it when discussing the history of racism and slavery in America:

“In exchange for white-skin privilege, the Euro-American workers accepted white identity and became active agents in the brutal oppression of African American laborers.

But they also fundamentally degraded their own conditions of existence.”

In other words, Haider contends that white supremacy is ultimately bad for white workers and thus all workers have an interest in abolishing it, whether they do or do not recognise that interest. What I want you to consider is that this might be the real hinge of the whole argument over identity politics, insomuch as this is a useful debate. Do you accept that the best strategy for the whole proletariat, people of colour and whites alike, is to oppose racism, or do you believe that the objective interests of the white section of the proletariat lie in supporting and upholding white supremacy? Identity politics comes from accepting the latter view. It takes both leftwing forms (moralism about race, hysterical demands that people must “renounce” their advantages etc.) and rightwing forms (fascism etc.) They are both two sides of the same coin insomuch as they both accept that white workers win from racism, they simply differ in their moral assessment of whether this is good or bad.

1. Cookies

There’s a sort of cognitive trap here that it’s easy to fall into. It’s easy to think that obviously all whites have an objective interest in supporting white supremacy. After all, by definition, white supremacy is a situation in which whites are better off than non-whites.

To see why this doesn’t necessarily follow, consider a cute little meme that goes around the internet sometimes, usually an image or text that is a variation on the following: “Bob gives Luke two cookies, Samantha one cookie, and keeps twenty seven cookies for himself. He then turns to Luke and says “Watch out! Samantha is going to take your extra cookie!” (For clarity, this is not a quote from Haider)

Bob represents the largely white bourgeoisie; Luke represents the white portion of the proletariat; and Samantha represents proletarians of color. It is simultaneously true that Luke is better off than Samantha because of white supremacy AND that both Luke and Samantha would be better off were white supremacy dismantled. Luke is better off than Samantha under this unjust distribution of cookies, but would still benefit from, and be

better off under, a more equitable distribution of cookies. To avoid confusion—this isn't just about money. Cookies also represent power, status, security and other goods.

Whether in relation to material things or more intangible goods, a system that guarantees supremacy over another group need not guarantee a genuine improvement in living standards overall.

2. Racism as a winning proposition for the capitalist

What does the capitalist gain from racism?

Our first clue should be that the most open and vicious racists are also the most open and vicious in opposing workers rights. The struggles are linked if for no other reason than the clearest opposition to both is one and the same.

Racism is used by the right (yes, even the respectable right) to constitute an alternative nexus of political struggle. What I mean by this is that politics is always a struggle between forces, but even the nature of this struggle is itself subject to conflict. The right benefits from conceptualising the fundamental political struggle as a struggle between races and nations whereas the left benefits from conceptualising that struggle as a struggle between classes, in which racism is used as a crucial weapon by the ruling class.

Someone who sees politics as fundamentally a struggle between races will engage in cross class collaboration for the perceived advantage of their race. They may vote for those they see as having linked racial interests, but who do not share their economic interests, such as white supremacist capitalists, or rich persons of colour with bourgeoisie values. They may also be reluctant to collaborate on issues that should unite workers, such as joining a union that includes workers of color. Another example would be contemporary white workers who refuse to oppose police violence, despite police violence being a threat to workers everywhere, because they identify the police with

their racial interests.

Thus racism creates a defanged and disorganised working class that doesn't even conceive of itself as sharing interests. It is a win for capitalists.

3. Not perceiving a shared interest is the root of many pathologies.

Arguably at least, The mistake of thinking that white supremacy isn't ultimately counter to the aims of the whole proletariat leads into all the other problems with identity politics.

For example, a common complaint about identity politics is that it is moralistic and preachy, and that this leads to a culture of infighting and vicious online arguments. This follows from believing that there is no common linkage of interests in the proletariat to oppose racism. If you truly believe that some workers benefit in the long-run from racism, you will naturally resort to preaching and moralism, since the white portion of the working class has no material interest in abolishing white supremacy, there will be no option but to prick their consciences to guilt. Because of the range of many different ways people's interests are thought to be irrevocably divided- (gender, race, sexuality, ability etc.) it's no wonder that a low trust, high nastiness environment develops. Such an approach is very different from classical Marxism, which posits that almost everyone (except a small ruling class) is united in their interests.

Another complaint that's frequently made about identity politics is that it is overly liberal, and not sufficiently committed to abolishing capitalism. Again, this pretty plainly flows from the view that the working class is not even potentially unified with respect to its objective interests on white supremacy. It's a pretty major blow to a Marxist anti-capitalist view of things if the supposedly universally emancipatory working class has no basis for a solidarity of shared interests in opposing racism. If the working class is nothing special in this regard, a space is opened for class collaborationism of a liberal form.

4. The shared political interest is not a pre-existing fact

As we talk here about common interests, please keep in mind an important caveat from Haider:

“A common interest is constituted by the composition of these multitudes into a group. This is a process of political practice.”

In other words, it is not so much that the working class already has a common interest in smashing both racism and capitalism, as that there is a potential liberating political strategy which could infuse it with a common interest and a common program. This is not a matter of a voluntary or subjective element deciding ex-nihilo to form such a coalition, rather it is a possibility already present in the class which comes to the forefront in certain circumstances.

5. Identity politics as a symptom of the constricted soul

If the root of identity politics is a rejection of the revolutionary anti-racist potential of the whole working class as a group with a common interest in abolishing white supremacy, where does it come from?

This quote from Stuart Hall reprinted in Chapter 5, discussing economic despair in the 1980's and its interaction with racism in the UK captures it:

“As economic circumstances tighten, so the competitive struggle between workers is increased, and a competition structured in terms of race or color distinctions has a great deal of mileage. It is precisely on this nerve that the National Front is playing at the moment, with considerable effect. So the crisis of the working class is reproduced, once again, through the structural mechanisms of racism, as a crisis within and between the working classes”

In other words, as Bob gets a larger and larger portion of the cookies, the idea that both Samantha and Luke's situation could get better at the same time seems more and more distant. Since Stuart Hall wrote this, the wage share has been in almost continual decline in the developed world—for over three decades now.

Crudely speaking these material realities come to be reflected in our souls, or as Haider more eloquently puts it:

“I have come to think that this sadness is the primary cause of the restriction of politics to one's personal identity. Not only has the idea of universal emancipation come to seem old-fashioned and outmoded, the very possibility of achieving anything beyond the temporary protection of individual comfort seems like a delusion. Hence a call for universally beneficial social change is often heard as a personal affront: instead of an affirmation of my individual demand for security and recognition, I am presented with a goal that lies beyond my powers to achieve.”

We need to move our imaginations beyond the equitable distribution of crumbs, towards a coherent anti-racist, anti-capitalist program that roots itself not merely upon an abstract notion of 'social justice', but upon a recognition of our common interest in dismantling white-supremacist capitalism. Moving our imagination in this way is supremely difficult, because our despair is not merely a voluntary choice, but an outcome of our circumstances.

There are however reasons to hope. It is difficult to read the political weather, nonetheless it seems to me that over the last two or three years, an understanding of the inseparable linkage between anti-racist and anti-capitalist struggle has continued to deepen, one sign of which is the publication of the book that we have discussed here. As both capitalism and white supremacy continue to be exposed and treated with greater cynicism there is space for us to draw the links, and an urgency for us to do so.

Seeing like a communist, an introduction to Marxism for an intelligent person who hasn't read on it previously

I read this:

“I picked up *Inventing The Future*... because I feel bad that I've never been able to get my head around the communist paradigm. In the past, I've learned new paradigms by reading a lot of books from within that paradigm (and hating them) and debating people from within that paradigm (and thinking they're crazy). Then fifty books and a hundred debates down the line, I finally get some kind of inkling of where they're coming from, and then after a while I can naturally make my mind shift into that mode and my only differences with them are at the high-level generators of disagreement. I was born into the Woke California Liberal paradigm, I managed to force myself to understand the libertarian paradigm in college, I managed to force myself to understand the right-wing paradigm a few years ago, and I would really like to be able to understand the communist paradigm too.”

And it got me thinking, is there some way that the process of grasping a *weltanschauung* could be sped-up? I could of course just write a debater's case for communism. However, such a case would just be another one of the fifty things that Scott would have to read before groking communism.

To grasp a *weltanschauung* is to rearrange what we always knew into another pattern. Once we've grasped it, it should seem surprising to us that we didn't see it. To that end, the easiest way to make the communist *weltanschauung* graspable seems to me to present a series of propositions that many would find at least plausible, which taken together gesture at another way of thinking. Rather than going for quality by trying to prove something really difficult and significant, let's go for quantity by presenting a series of fairly modest ideas that jointly gesture at something more.

I say 'gesture' rather than 'explain' because I suspect that when it comes to grasping a model of the world on this scale, part of the experience has got to be putting the pieces together for yourself. Some ideas always lose something if they are explained, to be properly conveyed they must instead be rediscovered by the student.

Perhaps the best model for this sort of thing is Wittgenstein's *Philosophical Investigations*. There's not a great deal of arguing going on there, but the ideas form a sort of haunting constellation. Or for another analogy, it's less a lecture, and more a kind of meandering stand up comedy set without the punchlines "what's the deal with language games eh?" It was after all Wittgenstein who said a serious and good philosophical work could be written that would consist entirely of jokes.

In what follows, I try to do that for communism, arranging a sort of blancmange of ideas that don't so much argue for a communist world view as swarm into it. It would be an interesting exercise to perform a similar conjuration for other world views, and if anyone gives it a try, please link me to it.

1. Method and social epistemology

1.1 The ideas on the table are a reflection of the kind of society we live in. The ideas that were on the table in AD 1019 are very different to the ideas on the table today. This is due to the interests of the powerful, but also due to the sheer difficulty of imagining your way out of your own society.

1.2 In all areas, but especially the areas that really, truly matter to the powerful (predominantly economic policy and foreign relations), the dominant ideas will disproportionately reflect the interests of the powerful, although of course there will be counter-narratives.

1.3 Even if all you want is an accurate view of the world, you need to be aware of this skew in the evidence that is being presented to you- a skew towards power. You need to work to overcome this skew.

1.4 Because power isn't absolute, the dominance of the ruling ideas is never absolute, and power comes to reluctantly compromise on matters that were previously off the table. For example, the discussion of universal healthcare in the US used to be regarded as a fringe position, but though the policy hasn't been adopted, room has been made at the table to, at least, discuss it.

1.5 You are not intrinsically smarter than a medieval scholar arguing that the great chain of being validates the divine right of kings. Don't think you can't be duped by ideas that will one day seem laughable. If those ideas favour one set of interests over another, the odds are that they will favour the more powerful interests.

1.6 The pervasiveness of ideology leads to a kind of circularity, or unseeing. Consider the argument 'capitalism is good because it's built on voluntary exchanges'. Remember that capitalism is only built on voluntary exchanges insofar as you accept that the people who society presently regards as the legitimate owners of goods really are the legitimate owners. Suppose that you didn't accept this. Suppose, for example, that you held that a family who has lived in a house for generations has a better claim to own it than the landlord. In that case it would seem to you that there is nothing voluntary about their tenancy- rather they are being extorted for something that should be theirs by right. Thus we come to the absurdity of the slogan 'capitalism is voluntary' it only works if you believe the capitalist distribution of property is right, but this is the whole of the topic under dispute! Ideology makes us engage in this kind of circular thinking, because it becomes tricky to suspend our belief in the existing system even long enough to argue for it!

1.7 Is communism primarily a theory about how things should be, about what should be done, or about how the world is? All three, in a difficult to describe and uneasy tension. They sort of loop into each other, and you could start that loop at different points. For example, we start a dream for how things should be, we ask why did that dream arise, and why is it frustrated, and so we come to ponder how things are at the moment, then the combination of a goal with a theory of the present conditions leads to thinking about a method to get to that goal

2. The material organization of society

2.1 Resources have always granted power and power has always granted resources, in turn, power and resources have always structured society. There are tight links between social structure, the interests of those who are in power, and the way a society produces and distributes resources.

2.2 Capitalism is a system under which people make investments for the aim of maximising return, with the return being reinvested (at least in part) into making more money, and so on. It can be seen as a growing spiral.

2.3 The process isn't entirely a matter of free choice. Rather, something like evolution is in play- if you don't try to maximise your returns, in the worst case you'll go bankrupt, and in the best case you'll become increasingly irrelevant. As we will discuss later, it starts to seem like capital has developed an existence, motivation and cunning all of its own, not fully reducible to what is in the head of any one person.

2.4 The pseudo-agency and intelligence of capital might fruitfully be compared to the pseudo-agency and intelligence of natural selection- a dumb process that, through the discipline of selection, thrives.

2.5 Under capitalism, capital is the motor of production. As a result, capital begins to discipline everything, because if you get in the way of smooth production, you're getting

in the way of a lot of very powerful interests. Any social institution- formal or informal- can thus be potentially subject to the discipline of capital. Historians have argued that things and concepts as important and diverse as religion, education, the family, and of course, politics and the state, have been relentlessly altered and rewoven by capitalism.

3. Pumping intuitions about fairness

3.1 Above we made the point that the voluntary status of capitalism depends on the legitimacy of a certain framework of property relations. Let's consider what the world looks like sans that framework. People often complain that chief executive officers make a huge amount disproportionate to the work they put in- but wait till you hear about shareholders! Under capitalism, extra money is owed to certain individuals in view of their 'ownership' of certain assets, and this starts to look like a kind of unearned privilege backed up by state power. That is- if you don't buy the legitimating narrative that the ownership of the productive assets is rightful.

3.2 The defender of capitalism will often try to argue that the order of property ownership under capitalism is legitimate and not arbitrary- it reflects a sequence of voluntary exchanges following legitimate initial acquisition of property from nature. There are two main sorts of problems with this. A) Technical philosophical problems about the Lockean or Nozickean account of distributive justice, which we won't go into here, and B) a simpler and arguably more decisive problem. History did not happen like that. It's simply not true that if you go back through the chains of ownership and exchange this is what you will find. It's war, theft and the state all the way down. The Lockean and Nozickean accounts of distributive justice have a peculiar fragility to them- it doesn't take much to render a chain of acquisition illegitimate, and there is much, from state subsidies to genocidal land-grabs.

4. Human Nature, hierarchy and markets

4.1 If we take human behaviour in the evolutionary environment as definitive of 'human nature', then human nature is not just passively egalitarian, but actively anti-hierarchical. Hunter-gatherers tend to actively suppress those who might otherwise become dominant.

4.2 The market economy is not atemporal. Other kinds of economies have existed. The notion that the market economy is 'human nature' would be very surprising to the vast majority of human societies that have ever existed (hunter gatherers).

4.3 But is it always the case that 'naturally' arising non-market based societies are too poor for trade to be very important? Hunter gatherers, medieval subsistence farmers and the like? No. Consider for example the Inca, who had a complex system of production managed through a 'naturally' occurring non-market economy without money or currency, in which households were issued with the goods they required from storehouses.

5. Technologically advanced non-market economies

5.1 Some non-capitalist economies existed relatively recently and using modern technologies. It is far from obvious that such economies were, in economic terms, failures. You may rightly doubt whether the Soviet Union is a good model for a communist society, but it is a very different model to contemporary capitalism, and its productive achievements were often impressive. This is somewhat odd because to listen to some people talk you'd think that societies not dominated by markets couldn't exist, but didn't some exist just thirty years ago? Everything that is actual is possible (and a great deal more than this).

5.2 But if a completely different kind of economy to our own can fly, why assume the ideal looks anything like this amalgamation we inhabit? Since Soviet style command

economies were viable, even superior in some ways, and among capitalist countries the variation in economic structure is vast, the space of possible ways an economy can be designed is large, and models quite different to our own can work quite well. That should shake your confidence- at least a little bit- that the economy you live in is optimal.

6. A Note About Weighing Costs

6.1 Never forget: we ignore the atrocities that are going on all the time around us because we regard them as a default state of affairs.

7. The Dialectic of Use and Exchange Value, and related social tangles

7.1 Our society produces things for exchange, not use, but this weirdly indirect way of servicing our needs has its downfalls. Producing things for exchange, not use, can fail to maximise utility for many reasons. Two of the most obvious are inequality (a person might desperately need a good, but not be able to afford it) and externalities (two people might happily buy and sell a good, without consideration for the positive or negative effects that seemingly private contract has on others. This will tend to lead to overproduction or underproduction of goods with strong externalities, relative to the social optimum).

7.2 Consider also the effects that producing for exchange, not use, has on the producer. Anyone who has worked as a telemarketer and gone home wishing they could do something that matters has felt this. There's a lot of jobs that just don't need doing.

7.3 Let's double back to ground we've already covered briefly- the eerie seeming 'self-moving' agency of capital. The operation of individual behaviour in a society with a given institutional framework gives rise to emergent phenomena that may have been

intended or desired by none of the participants, or by only a very few of them. Capitalism has a logic all of its own, that logic arises from our individual actions, but ends up controlling them.

7.4 This self-moving logic of capitalism will always work to subvert principles, legal and moral. Every institution faces the imperative 'serve the goal of profit', and everything and everyone that exists in the social world is warped by it.

7.5 But aren't these goals just human goals- weighted by purchasing power? Weighting by purchasing power would be bad enough, given that it makes the desires of some hundreds of millions of times more important than those of others, but the problem is deeper than this. The self-expanding loop of capital doesn't really care much about externalities.

7.6 As Gerry Cohen pointed out, it also has a peculiar preference structure regarding leisure- it would always prefer societies to work more and consume more, rather than use additional productive power for leisure. Why? Because capital owners would prefer for their capital to be employed for as large a portion of the day as possible, so as to maximise profits (it's not as if the capital owner has to be working all the time their capital is being put to work). It will push for this through means cultural and political- although it won't always get its way.

8. Planning? Yes, but.

8.1 Planning a better world is important, but don't mistake the absence of complete plans for the futility of transition. No one planned capitalism out of feudalism, but struggle and dreaming helped get us there. Plans are necessary, but they will never be complete. (Editorial note here: Showing my market socialist sympathies I recommend "Economics of a Feasible Socialism" if you are interested in detailed plans for a socialist society.)

9. Intentionally or not, people are lying to you

9.1 It's a childish delusion that you just so happen to live in the only civilisation without propaganda.

9.2 Propaganda is rife. Consider public debate about just about any policy position. You're constantly being told that even the slightest steps towards the subordination of exchange value to use value (e.g., universal healthcare) will maybe cause the economy to keel over dead and definitely rip ragged human economic activity in the sphere in question.

9.3 In many cases you know for a fact this can't be true, because even in the capitalist world there are many countries where any policy that might be under debate has already been adopted. In the healthcare debate, people will tell you that the economy will suffocate, or at least that healthcare will bloat and become ineffective, if universal healthcare is implemented, even though anyone can drive to Canada.

9.4. In the US minimum wage debate, people will tell you that unemployment would spiral if minimum wages were raised to \$15 dollars an hour. Meanwhile, in thoroughly capitalist Australia, a 21 year old fast food worker in Australia who is casual (without guaranteed hours) is entitled to \$26 an hour (and no, Australian dollars don't go much farther, or much less far, than US dollars). At every turn, capitalism is presented as at once essential to human activity, but also very fragile and in need of the velvet glove treatment. This isn't true- capitalism can be configured in many different ways and still work, just buy a plane ticket if you don't believe me.

10. The State

10.1 On no topic is the communist at more danger of being misunderstood than the state. It didn't use to be this way- in the past it was fairly easy to understand that the

existing state is no friend to the communist, but with the advent of the post WWII welfare state, people began to associate the state with communism.

10.2 From the communist viewpoint, the welfare state can be seen as a sort of commons that has been won as a concession from the ruling class- an exception to the usually strictly individualistic property order the state typically administers. Because communists do not regard capitalism as natural, for the communist the welfare state is not an unnatural 'exception' to the general order of things, it's just one more way the state organises property rights, a steam valve to prevent the existing state's real purpose- the maintenance and improvement of private property- from exploding.

10.3 There are three factors that act on the state to pull it towards the interests of capital. The first is the least significant, but the most noticed- it costs money to run an election campaign, and money can buy lobbying. The second is that the most important people in the state, politicians and senior civil servants, tend to be rich, have rich friends and have gone to the same cliques of schools, universities and jobs as other rich people, forming a well connected power-elite. The third is the power of the purse strings- states are disciplined in supporting capitalism through the threat of capital strike and diminishment in investment.

10.4 This critique of the state can go in two directions for the communist. One either thinks that the state can be done away with entirely (anarcho-communism), or, regarding the state as a body of armed people for the defence of property relations, one concludes it is necessary to replace the capitalist state with a communist state (main-line communism). Optimists in the latter camp might hope that the need for such a state might one day wither away, a more pessimistic view (which I favour) is that complex societies probably inherently require standing armed bodies for the enforcement of social order.

11. Workers: Is there a protagonist to lead us through these brambles?

11.1 How interesting that the very people who have the least stake in the system (wage labourers) are also the very same people without whom the system would grind to a halt!

11.2 To be more explicit, think about things like a murder case in reverse. If capitalism is going to be killed, the question is, who has means and motive. From this point of view, proletarians- people who sell their labour for a living- would seem like the perfect suspects for the pre-crime of murdering capitalism. They gain the least from capitalism, yet oddly they are absolutely essential to operating it. Proletarians have means and motive. The method? Ah, well that's a very hard question.

Appendix: What I actually think

Having outlined a worldview, I should probably not be a coward, and state plainly what I think. The truth is though that when it comes to ultimate political questions I'm deeply uncertain.

I believe that in an ideal society, production would be much more closely regulated by what people need rather than for exchange. Markets in this regard are a bit like fire- a potentially useful servant, yet a terrible master.

I don't have any blueprints for the future. Not because I have some principled objection to blueprints, but because I simply don't know.

I'm confident that a more equal and more democratic society is desirable. I'm confident that a combination of both political and technological advances will be essential in getting us there. I'm confident that we can do better than this. How much better we can do, and how? They're difficult questions.

Of fake fire-starters and real panic-mongers

As the Paris Commune was ending, there was a great fear that the city was filled with crazed women who, out of pure bitterness and spite, were roaming the streets, and setting buildings alight- the so called Pétroleuses.

Modern historians regard this as a fabrication, but at the time, France was gripped with the idea of such people, who represented fears of political excess, the lower classes, and of course, women. The nature of fire itself encourages such paranoia- its effects are much more visible than its causes, allowing great room for speculation. The history of attributing arson to political enemies is well documented and most famously includes the Reichstag fire. Mysterious fires have often been falsely attributed not just to political agitators, but to cultural ethnic and religious scapegoats, like the Jews.

In Australia during the 2019/2020 fire season things have gotten bad. Vast tracts of land have been burnt, more so than in any single previous fire season. The overwhelming consensus among scientists ranging from meteorologists to fire behaviourists is that this is, in large part, a result of climate change.

Naturally this is awkward for those who passionately reject the existence of climate change. For the terrible injustice of being required to explain something inconvenient to their world view, many denialists decided it would not be enough to simply refute the claims, rather, they must find some way in which the fires were caused by their mortal enemies, the environmentalists.

Deniers first tried to argue that the Greens had caused the fires by opposing hazard reduction burning. This was shown to be an obvious lie- the Greens have never held government anywhere, people involved in hazard reduction burns say it isn't true, and so on.

Their next narrative, now emerging, also conveniently places the blame on their enemies. Environmentalists arsonists are deliberately lighting fires in order to create fear of global warming.

As global warming gets worse, we should expect to see more attempts by the right to shift the blame for the effects of warming onto the left, and other enemies of the right such as ethnic minorities. Fires will be blamed on environmentalists and Islamic terrorism. Food shortages will be blamed on left-wing conspiracies. Power outages will be blamed either on progressive policies, or on sabotage. Steel yourself for it now.

Because conspiracies and denialism are unaccountable to reality, they can turn any event into an expression of lurid fears or dreams, a political weapon, or all of these. As it becomes more and more unmoored from evidence, denialism becomes a mental playground for petty hatreds in which the believer can play out fantasies about the wretched evil of their enemies. Various frustrations and fears- whether about ongoing catastrophes, or just about the life of the fantasist, can be displaced onto the hated foe.

In this fantasy world, we see an inversion. Minorities and the politically disenfranchised are no longer weak- rather they are strong, strong enough to carry out malicious and global schemes. This inversion enables hatred- only a handful of utter psychopaths will admit, even to themselves, that they hate the weak, so it is necessary to paint the weak as strong. It's an upside down ghost-world, but its popularity shows it must make a kind of emotional sense.

What strange creatures we are, to find pleasure in hate.

Philosophy

Meeting Nietzsche at the limits of rationality and the limits of Analytic Philosophy

Note: I am not a Nietzsche Scholar. These are my musings and attempts to form an internal model of something. It is my hope that, even if they are wrong, they are wrong in an illustrative way. In this regard, let me appeal to Nietzsche himself: "It is certainly not the least charm of a theory that it is refutable; it is precisely thereby that it attracts the more subtle minds."

I've been reading Nietzsche lately. His genius is pretty clear to see, though on most things we disagree. My reading style is a little naive. I've picked up copies of his books, and I'm going through them by themselves, rather than reading them with secondary sources. So far I've finished *Beyond Good & Evil*, *The Birth of Tragedy*, *Thus Spake Zarathustra* and hope to go through more.

It's got me thinking about the limits of Analytic Philosophy. Nietzsche's own contempt of threads that would later weave together into Analytic Philosophy- British Empiricism, British Utilitarianism, and (what was then) Continental Positivism is quite clear. Still, I thought it might be productive to mash some ideas from Nietzsche into a constructive critique of the Analytic tradition.

Again, this is a very strange sort of conversation I am orchestrating. I am reading Nietzsche, as it were, with apparent innocence. Perhaps I'll be criticized for going outside my sub-sub-specialty, but I figure that when a philosopher publishes a book, even if it's a demanding tome like *Beyond Good and Evil*, they entitle anyone willing to honestly read it to take a punt at it.

Being creative and understanding creativity: Into the hidden chamber of discovery

Nietzsche is, in a way, sloppy. He contradicts himself a lot. He doesn't have a consistent line on questions like Is the truth unobtainable, or is obtaining it dangerous? He thinks that the origins of a thinker's thoughts is a far more telling objection against those thoughts than it really is. Regardless, he has a lot of insights. I am told that he is not the only guy like this in the continental tradition (Fn: Which is not to generalize all continental philosophy as sloppy). Foucault, for example, is sloppy on both history and philosophy, according to philosophers and historians I trust who have read him. Yet, when read in moderation and with a skeptical eye, Foucault offers insights.

In relation to creative endeavors, the argument that sloppiness can be useful isn't controversial. Someone on acid can't reason as well as a sober person, but there may be many forms of creative work for which they are better suited. LSD was instrumental in creating much brilliant music and poetry for example. In such cases, the impairment of rational thought is glued to the rise of creative powers. I see no reason to think this phenomenon is exclusive to the arts. There may well be true or useful characterizations of the world that are easier to find if rational thought is loosened.

Even if you dispute the insights of Nietzsche, he isn't the only possible example. Marxism is a great case study. Analytic Marxism was an attempt to translate Marxism from the continental style into the reasoning style of Analytic Philosophy. It's a cool school of thought, but it's almost impossible to imagine a world in which Marxism was first invented within Analytic Philosophy. Not by accident did Marxism grow on the altogether wilder and more speculative grounds of Left-Hegelian philosophy.

Nor is this phenomenon exclusive to philosophical ideas like Marxism. Many of the great physicists of the early 20th century- think Schrodinger, Einstein, Heisenberg, etc., drew on philosophical ideas that we can politely call "ripe with speculation". It is

impossible to say to what extent such speculative ideas were actually useful in the scientific thought of these men. Certainly, though, these physicists themselves thought these ideas were useful.

So I suspect the sloppiness and the leaps in logic are active ingredients in Nietzsche's work, allowing him to obtain his biting insights.

If this is right, we face a difficulty. How should we trade off a particular kind of creativity against a particular kind of rigor? How are we to evaluate methodology, if the truth or reasonableness of methodological premises isn't the only factor in evaluation, at least sometimes, and in certain kinds of inquiry?

The best framework I've been able to come up with so far for understanding, though not solving, this impasse is the distinction between The Contexts of Discovery and Contexts of Justification from the philosophy of science. The context of discovery is the context of creating new good "guesses" about the world (e.g. hypotheses). The context of justification is about justifying- or disproving- those hypotheses.

Analytic philosophy has made great strides in understanding not only deductive reasoning but also inductive and abductive reasoning for and against existing views. Evaluating, understanding, and encouraging the creation of new ideas though, remains largely beyond its reach. In other words, Analytic philosophy has proven much better at understanding the process of justification than it has at understanding the process of discovery- through some are trying to remedy this.

Not only has Analytic philosophy been better at understanding justification than at understanding discovery, but it's also arguably been better at doing justification (and refutation) than the discovery of new positions. This is most true in relation to social philosophy and theory, and in relation to philosophical psychology. This is by no means

to downplay the brilliant new positions that have been created by Analytic Philosophy, enough to fill libraries, but it is something to think about.

By contrast, not only does Nietzsche excel at finding novel hypotheses, he is very much at home theorizing about the context of discovery and ushers us into that place.

Nietzsche excels in generating novel and interesting- if perhaps not always persuasive- theories of the genealogies/origins of ideas and institutions. Even more so, reading his own work, we are left with a sense of philosophy in motion. Often reading, *Beyond Good and Evil*, it seemed I was reading thinking, rather than reading the products of thought- philosophizing, rather than philosophy. There is a kind of teaching about the creative process through demonstration to which he treats us.

So what is the way forward? As I mentioned earlier, evaluating scholarly methodology becomes monstrously harder when it's not a matter of 'just' looking for more truth and rigor in the methodological postulates. I don't have any sense of the right way to balance rigor and soaring when it comes to thinking. My only thought is that one virtue of intellectual diversity is that at least no individual needs to do it all at once.

The human as a believer and the philosophy of belief

I am no historian of thought, but it seems to me that the classical picture of humans in philosophy and "western" thought is that people are believers. we believe things, i.e. we have a stock of things that we hold to be true. At least until the modern period, the status of desire in this picture is unclear. Maybe desires are separate things from beliefs, or maybe to desire X is to believe X is excellent or worth possessing. Regardless, belief takes the lead, especially in premodern philosophy which tends to disdain "the passions". Exactly what the passions are- whether they consist in all motivations, all desires, all emotions- what is the relation between these in turn, etc. etc.- is a little unclear. Nonetheless, the passions are certainly distrusted.

Hume comes along and makes the argument that reason is the slave of the passions. What he means by this is that no amount of reasoning can ever lead you on its own to want something. In practice, Hume means by this means desire is autonomous from belief. No belief/theory/conception can ever imply a desire/passion/motivation and vice-versa. [Aside: this is very closely conceptually linked to Hume's further claim that one cannot derive an is from an ought and vice-versa.]

So we get a bifurcation- there are beliefs and desires. In more modern times, formal models of mind and agency have become increasingly sophisticated. There are degrees of belief over different possible states of the world, and utility functions over these same states. Belief still retains, at least in the context of our thinking about philosophy, a kind of thematic primacy.

I previously have worried a lot about one aspect of this account- the idea that people have clear beliefs. Rather- I think that there are many different components to what we call belief. Often these components come apart- so you can believe in one thing in one sense of belief, and disbelieve it in another sense. Tamar Gendler was one of the first to pioneer this "splitting" of beliefs into multiple components with her concept of Alief, but I think there are many other components as well- at least four by my count.

Nietzsche, I think, turns this picture on its head in a different way. He imagines a philosophy in which the central object is not what we believe, but what we desire and will.

These two critiques- the one I like to push, that belief is fragmentary, and the one Nietzsche likes to push, that our desire is often more central to who we are than our beliefs, complement each other. In many ways, what we desire in our lives and world may be much more stable between contexts, and much less fragmentary, than what we believe about our lives and the world. To the extent that there is any continuity or

wholeness to a person whatsoever, it is much more in what they want than in their internal map of the world, still less the stories they tell about why they are doing things.

There is an interesting analogy here, I think, in inverting the usual priority of belief and desire, with the Marxist inversion of Hegel, in which material circumstances and the means of subsistence are seen as having precedence over ideas and ideology. Beliefs stand for ideas, desire for the material basis of life. But that's a topic for another essay. I'll need to do more research because a lot of people have probably already said it.

Jumping ships on moral reasoning

The typical image of moral reasoning in Analytic Philosophy is some variation of the following. Since one cannot derive an ought from an is, one gathers together all of one's oughts- that is all one's moral urges. Some of these will be contradictory, one seeks to make adjustments to bring them into line with each other, smoothing out real and potential contradictions, and creating a coherent system of moral requirements. This is the method of reflective equilibrium. The philosopher works the jagged clay of moral intuition into a manageable system. I tend to think that even many Analytic Philosophers who claim not to be following this approach covertly are (much to Kieran's dismay).

Although the term reflective equilibrium did not exist at the time, Nietzsche makes it clear that he is unhappy with processes like this. The philosopher shouldn't merely smooth over existing values- they should be a creator of values, or at the very least, an evaluator of values.

But is this possible? Nothing comes from nothing. The only argument in favor of a value can be another value since one cannot derive an ought from an is. The problem, then, with "creating" values without reference to existing values is that we have nowhere to stand.

Otto Neurath, in speaking about beliefs rather than moral principles, said something that can be applied to morals:

“We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this, the rest of the ship is used as support. In this way, by using the old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction.”

But there is, I think, one sense in which we can critically consider our values in a way that goes beyond reflective equilibrium.

Let's say you were to sort through your moral views and come to some rough sense of what you stood for. It's possible that looking out from that vantage, you might decide that another set of values is practically dominant. By this I mean that adopting that set of values, genuinely committing yourself to them, will make you more likely to maximize not only your own current values and goals but also the new values and goals you adopt. The new set of values might be the best option both from the point of view of the new values and from your old set of values.

This is a bit abstract, so let me give an example. When I was severely depressed, I found myself contemplating what I stood for. At the time I saw justice as particularly central to my basic, bedrock morality- justice as an important end in itself. I realized though, that this overweening concern for justice was part of what had made me sick and depressed. By leaping from the values I held then, to a system of values based more on compassion, I would be more able to do good both from the point of view of my old, justice-centered values and from the point of view of my new compassion centered values. This practical dominance argument induced me to change. Through a process that was both practical and philosophical, I rejected an old set of values and adopted a new one.

This, I think, is one way in which Nietzsche's ideal of the philosopher as an evaluator and creator of values- and not merely a smoother and summarizer of them- is possible, but without requiring us to step to an archimedean point outside all value.

[Aside: I do not think this situation I have described is especially rare. It seems to me quite common that two systems of values have mostly overlapping goals, but that one system could be better at getting at those goals than the other. Here's another way it could happen. You might find that one system of values is very vulnerable to being rationalized- manipulated to get the result you want- due to the intricacy of its fine distinctions and casuistry. This makes you think you should prefer another code that, in practice, overlaps in its ultimate ends to a large degree with your current code, but is less vulnerable to being rationalized in this sense. Thus, you switch.]

The neglected emotional-aesthetic questions of philosophy

Nietzsche draws our attention to what we might think of as emotional-philosophical problems. Problems that are, to speak somewhat inaccurately, non-cognitive.

When I was in high school my teacher, in talking about our essays on King Lear, suggested that we had to grapple with the problem of evil. I was a little perplexed. I was not religious, nor, to my knowledge, was she. What is the problem of evil in a world without God? There exists evil. It's sad. The end.

I still don't know what she meant by posing this, perhaps she didn't mean anything very definite. Having read Nietzsche though, I think I see both the outline of a secular continuation of the problem of evil, and a secular solution.

Nietzsche grapples with Schopenhauer's pessimism. In a very crude sketch, this pessimism goes as follows. Life is a struggle. We seek what we desire. Satisfaction is only a brief respite, leading to the next round of seeking and frustrated desire. The

overall picture given by Schopenhauer is very nearly identical to that propounded by the Buddha. To exist as a being is to want things. To want things is to be frustrated and incomplete.

The new, secular, problem of evil then is this, how can we give meaning to this process in a way which makes the continual frenzy of frustrated desire acceptable?

This problem is what I call an emotional problem. It's a problem with the aesthetics of life. How are we to understand this overall picture in a way which makes it less horrifying, or, if we cannot do that, how are we to resign ourselves to the horror? It's not a factual question about the way things are. It's not even really a moral question about how we should live. It's a question about how we can keep perceiving life as beautiful and enjoyable.

This kind of aesthetic or emotional problem- of making sense of patterns in a way that allows us to process them or cope with them-, is a lot of what ordinary people mean by philosophy. Sadly, it's a pursuit that's often missing from Analytic Philosophy, because what is being searched for is not so much a hidden truth or even imperative- but a method of setting life to the right kind of narrative.

Nietzsche's solution to the problem of pessimism is that the struggle itself has to be seen as valuable- and not merely success in those struggles. By learning to love the beauty in the struggle to live well, we can reconcile ourselves to the permanent hunger of existence. This is a way of framing the aesthetics of life which makes it something other than ugly, and only cruel. I'm not sure if this is quite the solution- but I think it's on the right track.

[Aside: I would add, though I don't know that Nietzsche would agree, that the view that certain kinds of striving against obstacles, might have intrinsic value is no reason to keep brutality, crippling diseases, hunger, etc. Even a life in a utopia of material comfort

has forms of striving enough- e.g. for artistic or scholarly greatness, for love, to be a good person, etc. I would rather a world in which people fight for self-actualization in Maslow's hierarchy than struggle for food and shelter. This is really a strong argument against social Darwinism- we must give people space to struggle with the higher and more beautiful problems.

I would add also that none of this is broadly opposed to an account of ethics focused on welfare- so long as welfare is understood broadly.]

This secular problem of evil is just one emotional problem in philosophy, there are assuredly many others. It is right to expect attempts at solutions to such emotional problems from philosophers, and it must be seen as a limitation in the analytic tradition that it has not, hitherto, often attempted to provide them.

In the not too distant future- things which might have been only a bare hint in Nietzsche's day- like the possibility of neurologically eliminating thwarted wanting and feeling altogether, might transpire. In a world in which "wire-heading" is possible- the question of the value of striving and thwarted desire may soon become urgent.

Four parts of belief

A little bit of prodding suggests that beliefs are not so simple as they seem. Consider for example Tamar Gendler's concept of an Aelief— a kind of belief-like state. An Aelief, per Wikipedia is:

“...an automatic or habitual belief-like attitude, particularly one that is in tension with a person's explicit beliefs.

For example, a person standing on a transparent balcony may believe that they are safe, but alieve that they are in danger..”

Of course there are other ways of dividing things up. When I was a wee undergraduate philosopher my lecturer gave the following case study. Young Catholic men claim to believe that the sin of self-abuse risks their immortal souls. Yet they engage in it with great enthusiasm. Traditional accounts of belief and rationality have difficulty making sense of this. There are, my lecturer suggested, three possible explanations:

- A) They don't really believe that self-abuse will send them to hell (though they believe that they believe this.)
- B) They are acting incredibly irrationally.

Or, his preferred option C) Belief is not a single thing. It consists in a complex of behaviours, thoughts and feelings which can be separated out from each other. In this case, two parts of what belief normally is -acting consistently with a view that X and sincerely asserting that X—come apart.

To speculate a little further, the brain is made up of a bundle of systems which are not as

well integrated as we might imagine. Thus it is at least possible that the seeming hypocrisy of these young catholic men arises from different mental systems having different models of the world.

I am reminded of the two streams hypothesis— viz, that there is evidence of strong segregation between the brain pathway involved in visual awareness of our environment and the brain pathway involved in visual action guiding through our environment.

Here is a list of the separable components of belief I'm aware of. See if you can think of any others:

1. The non-verbal action component

If I believe it is raining outside I instinctively grab an umbrella on the way out. If I believe the price of oil will fall tomorrow I won't buy oil now. An important subcategory here is betting behaviour, it may lie somewhere between 1&2.

2. The sincere assertion component

This component of belief is the ability to assert with real sincerity that P is true or that you believe P to be true. As in the case of the young catholic men we described above, it can come apart from the non-verbal action component—especially in matters of sacred belief.

3. The feeling component

Often, someone with an anxiety disorder can assert that something they fear is not true. Their non-verbal actions may also reflect this apparent disbelief, to varying degrees. Yet they are still very distressed by some troubling possibility as if they believed it were true or going to come true. We might call this the feeling component of belief. This is

primarily what the concept of aeliefs gets at.

4. The commitment component

This one (owing to Kieran Latty) is a bit less well defined than the others. It consists in a determination to believe X, in the sense one or more of the modalities of belief listed above. Suppose for example that I have no real conviction that humanity will survive the next hundred years, but I consider it strategically important that I believe such a thing. I might be committed to making myself act, sincerely speak and feel as if it were true that humanity will survive. Whether or not I succeed in this commitment, I have the commitment component of belief.

A sketch of a layered solution to the interpersonal comparison problem

“I can’t explain, the state that I’m in, the state of my heart, he was my best friend.”

-Sufjan Stevens, Wasps of the Palisades

My Ph.D. is partly about the interpersonal comparison problem. It’s a somewhat nerdy topic, often regarded as insoluble. After explaining it to you, I want to convince you that A) that it’s a really important question and that B) there is a solution that has been hiding in plain sight. What I also want to do is to write an essay on a technical philosophical problem in a very accessible way. There’s a great deal of mystery about what it is philosophers do, so I want to draw back the curtain- not by description but by example.

The “answer” takes a special form. I first propose a solution. I then say “well suppose you don’t accept this assumption in the solution, if you just accept this alternative, weaker assumption you can still get the result”. Then I weaken it again and so on.

It is my hope to show that, so long as you think that the branch of psychology known as psychometrics is broadly acceptable in its methodology- at least as provisional best practice- and you are willing to make some very weak additional assumptions, the interpersonal comparison problem is a solved one. Or rather, the interpersonal comparison problem is solved as a practical barrier to ethical inquiry that relies on interpersonal comparisons. There may still be lingering conceptual questions, but no one should hold these questions up as a reason not to use interpersonal comparisons in ethical inquiry- for example, in thinking about welfare economics.

But I’m getting ahead of myself. Here’s an example to introduce the problem we’re talking about. A hospital is running low on painkillers. There is only enough left for one

patient. One has a headache secondary to a head cold. The other has dislocated their shoulder. Which should get the painkiller? Most of us will have zero problems coming to a resolution, the patient with a dislocated shoulder should get it. But what does it mean for one person to be in greater pain than another? How can we quantify pain in a way that can be compared across persons?

Even in the form, I have put it, this is not a purely hypothetical problem. I have worked in reception jobs in hospitals and watched medical staff make decisions about triage etc. partly on the basis of considerations of the degree of pain between individuals. If there's no scientific way to make those judgments, that's bad news. If those judgments are meaningless even in principle, that's even worse.

What I'm speaking of is the problem of interpersonal comparison. It's not just a philosopher's question! In economics, at least since Lionel Robbin's book on the subject, it has been something of an article of faith that interpersonal comparison is difficult, or fraught with difficulties, or perhaps is nothing but the representation of a decisionmaker's own preferences over the tradeoff rate between people. On the basis of skepticism about interpersonal comparison, welfare economics became all about Pareto improvements and the Kaldor-Hicks criteria (if you don't know what these mean, don't worry, it won't matter here). This, in my opinion, contributed to the perception of economics as an anti-egalitarian science. It may have even led policy in an anti-egalitarian direction.

Practical applications

The problem of interpersonal comparison is a very general one. Consider the field of effective altruism, which aims to maximize the good done with a given set of resources. Good in this field is often conceived of in terms of human (or animal) welfare. Comparing two proposals in terms of their effects on human welfare almost necessarily requires quantifying mental states.

Here's why. Two of the most popular theories of what it means for a person's life to go well are the hedonic theory, according to which one's life goes well to the degree that there is a preponderance of pleasure over pain- and the preference satisfaction theory, according to which one's life goes well to the degree that one gets what one wants. On both these theories of welfare, quantifying the benefits of different welfare improving programs will likely require comparing the intensities of different mental states across different people- pleasures, pains and wants.

There is a third theory of human welfare- the objective list theory- according to which a person's welfare is constituted by the degree to which they have certain good things- like friendship, opportunities, security etc. This might seem to get us out of the problem of having to compare intensities of mental states, but really it doesn't, because, in almost all plausible versions of this account, pleasure and desire satisfaction are important items on that list.

So we can't really do effective altruism without some method- even if only an implicit one- of comparing the intensity of mental states between people.

Beyond effective altruism, consider also the problem of artificial intelligence alignment. Much research is happening at the moment on the question of how to define human ethical priorities formally in such a way that an intelligent machine could be instructed to respect them. On most accounts of ethics, part of our informal, everyday ethical calculus is making these interpersonal comparisons. A clearer understanding of how comparisons can be done in a principled way is thus necessary for AI alignment research.

Defining the problem

Attentive economists and philosophers might have noticed that I speak here of the interpersonal comparison problem. I do not speak of the interpersonal utility comparison problem which is its more common name. This is for two reasons.

The first is that utility is a poorly defined term. It is sometimes treated as synonymous with welfare or wellbeing and it is sometimes treated as synonymous with preference fulfillment (as in the Von-Neumann Morgenstern utility model). Even more confusingly, these two things- preference fulfillment and wellbeing- are sometimes treated as synonymous with each other and sometimes not.

The second is that there are interesting problems about comparing mental states that may not be directly related to utility at all. I might want to say “Bob is feeling angrier than Alice”, and although, of course, Bob’s degree of anger is related to both his utility and his welfare, on no definition whatsoever is it constitutive of it. The problem of how to compare Bob and Alice’s degree of anger, and the meaning of such comparisons, is an interesting problem in and of itself.

So what we’re really interested in comparing is the intensity of certain kinds of mental states between people. Exactly what is in this bundle of mental states is a little difficult to enumerate- but I would put forward, as a basic list:

Pleasures and pains

Desires and aversions

Emotions

There is no need to include beliefs at least on a certain definition of belief, as Bayesians have given us an adequate account of how to compare the strength of beliefs using betting behavior.

I call this category affective mental states because they all seem to have a tight conceptual link with motivation.

Empirical usefulness and psychometrics

Before certain complexities are added, I don't really think that the interpersonal comparison problem is that difficult. Consider, what makes us think that we can compare temperatures between objects? We develop hypotheses about ways to measure temperatures, and how hot and cold certain things are. We find that using these hypotheses we can do empirical work- make predictions and so on. That's really all it is.

Can a guess about the relative intensity of some affective mental state do empirical work? Can it help us make true predictions, and not lead us too often to false ones? Yes!

There's a whole science called psychometrics which makes estimates of the magnitude of various mental constructs, including, but not limited to, affective states of all the types we discussed above. Indeed, within psychometrics, there is a field of happiness studies, focused specifically on constructs like life satisfaction and happiness that many consider of one essence with welfare itself.

But these psychometric approaches were historically neglected by economists and philosophers working on the problem of interpersonal comparison. Indeed, psychometric approaches have often been neglected in general in these fields- though this is changing now- see the emerging field of happiness economics and the work of the philosopher Alexandrova.

There's a philosopher called Angner who has been working on the differences between psychologists and economists in the measurement of welfare for a while. His thesis is that it comes down to different understandings in the theory of measurement. Psychologists use a more flexible, one might say, empiricist, approach to measurement called the psychometric approach, whereas economists prefer the representational theory of measurement, a more rationalist approach which is based on formal axiomatizations.

The way psychometrics and the psychometric theory of measurement operates is by assigning magnitudes to a person's level of a construct through tests with standardised items ("Barry's level of happiness is 7/10 whereas Alice's level of happiness is 9/10") and then using those assignments of numbers to make predictions.

Let's say that we're measuring happiness. We begin by creating a series of questions that we think, based on our understanding of happiness, should measure happiness.

E.g., rate the following propositions 1 to 5, with 1 being strongly disagree and 5 being strongly agree.

I am generally in a positive mood.

I feel good about life.

I am a happy person

We can see that the test has a certain plausibility because its questions are conceptually related to happiness (it has "face validity"). Even this alone gives us some basis for credence in the test as a measure of happiness.

Now we administer it to a bank of people, using it to assign estimated happiness scores. We first check to make sure that it is measuring something and doesn't just represent random noise, this is to say we check the measure for reliability. There are a couple of different ways to do this, but one is to administer the same test to the same group of people with, say, an hour's gap between, and check the correlation between the scores at T1 & T2.

Our next task is to check how well it is performing as a happiness estimator. We might use its estimated happiness scores in a regression model to predict results on other tests which measure similar things (we call this “convergent validity”). For example, we might correlate it against a preexisting test of, say, hopefulness. One interesting form of convergent validity is to compare first-person results with third-person results. Have Bob fill out the test, then have Bob’s roommate, Alice, fill out the test as if she were Bob and see what the correlation is.

Or we might use test results to predict a behavioral outcome like suicide rate or frequency of smiling. This is criterion validity of a type we might call behavioral validity.

We might also flip things around and see how well circumstances, like an unhappy breakup, can predict our assignment of scores via a test. This would be another example of criterion validity- I call this subtype situational validity.

We might even develop hypotheses about how our measure should be related to biology if it really does capture happiness. For example, we might check to see if it is inversely related to stress response hormones like cortisol. This would be another type of criterion validity we can call biological validity.

Thus psychometrics gives us a way to estimate the relative intensity of affective states. It then tests these estimates, seeing if they are borne out in behavior, environment, biology, peer opinion, and other tests. Through an iterating process of testing, theory development, and application psychometrics aim at better and better ways of assigning numbers to mental states in a way that is valid across persons. There’s a lot I haven’t gotten into here, including more details on the role of statistics- especially factor analysis and psychometrics is not a field without methodological controversy, but, overall, it seems psychometricians never got the memo about the impossibility of interpersonal comparison.

The escape route

To me it seems that psychometrics is measuring something interpersonally between people- its capacity for empirical success shows this. Thus, a skeptic of interpersonal comparison owes us an account of what psychometrics is and is not measuring if they are to maintain that interpersonal comparison for ethical purposes is impossible.

The most plausible approach here is to insist that there is a distinction between mental states conceived of in how we experience them and conceived of in how they influence our behavior. This probably seems very abstract, so let me explain.

Consider the concept of qualia. To introduce the idea of qualia consider Alice. Alice has spent her whole life seeing the color spectrum inverted. Her greens are reds. Her yellows are blues.

However, from a young age, she was taught language like everyone else, to associate words with colors she saw. Thus she calls her green experience of what we consider to be a red object “red”, just as we do. Presumably, no one will ever even know that Alice’s experiences are so very different from ours in this way. This “greenness of green” is what we call qualia, and though it seems immediately present in consciousness, it’s hard to imagine what difference it could make to behavior.

It might be a short step from admitting qualia to making interpersonal comparison impossible. Consider the feel of desire, of longing. Now imagine that all your longings and aversions were exactly twice as great. You might think that this would have notable impacts on your behavior- perhaps making you a more passionate person- but there is a strong argument that it wouldn’t. For example, your strengthened desire to act might be exactly counterbalanced by your increased laziness. Perhaps then, experience can be altered without a functional alteration in the case of affective states, if you scale them in proportion to each other.

So, in order to prevent the conclusion that psychometrics can be used for interpersonal comparison, what our interlocutor is aiming at is a bifurcation between the functional part of an affective state (which we will call an f-state) and the experiential part of an affective state identifiable with qualia (which we will call an e-state). It is acknowledged that psychometrics can measure and compare f-states, but e-states are more mysterious and inscrutable- hence thwarting efforts at interpersonal comparison. Remember those terms, e-state, and f-state, they're going to keep coming up.

If you're wondering why psychometrics can only measure f-states, remember that what psychometrics measures is behavior (even if it's only question answering behavior), and that if it influences behavior, it's part of the f-state.

There's a further assumption here. The critic assumes that it's these scientifically inscrutable, interpersonally incomparable e-states that matter for ethical purposes- it's these states which comprise human welfare or suffering. If we acknowledged e-states existed but didn't regard them as ethically important, they wouldn't be troubling from the standpoint of ethical decision-making or policymaking. Thus while they would create difficulties for interpersonal comparison, they wouldn't be difficulties of practical relevance.

Cutting the escape route off at the pass: functionalism

In the previous section, I explained a way out of the seemingly obvious conclusion that psychometrics enables interpersonal comparison. That escape route was to disentangle feeling and behavior in a particular way. In this section, I'm going to outline a counterargument against this "escape route".

Functionalism is a view in the philosophy of mind about what the mind is. It might be best to explain it by way of comparison to analytical behaviorism because it can be seen as a more evolved version of that doctrine.

Analytic behaviorism, a now almost extinct view in the philosophy of mind, held as follows. Let's say you are angry. That anger is constituted by certain behaviors and behavioral tendencies. For example, you may raise your volume, tend to act destructively and rashly, become flushed in the face, etc. Those behaviors and behavioral tendencies are your anger. Analytical behaviorism has the advantage of being a purely physical view of what the mind is, but it has disadvantages. For example, we generally think that your anger causes you to raise your voice. But if your anger is partially constituted by your tendency to raise your voice, it's not really accurate to say that your anger has caused you to raise your voice.

The functionalist has a solution to these and many other problems of analytical behaviorism. What if your anger is whatever arrangement of your central nervous system that causes you to behave in an angry way? This keeps a tight conceptual connection between behavior and mental states while making sense of our ordinary intuition that mental states cause behavior.

Functionalism abolishes the possibility of a residual unobservable difference in mental states by holding that e-states separate from f-states don't exist.

There are many good arguments for this kind of functionalism that denies there are separate f-states and e-states. Consider, for example, that if e-states truly are separate from f-states, they have no influence on behavior since f-states can include anything that has an effect on behavior. The theory then faces a problem why are we talking about e-states if they have no influence on behavior? (For those interested, this objection mirrors a classical objection to epiphenomenalism in the philosophy of mind).

So if you accept functionalism, your confidence in psychometrics as a yardstick of interpersonal comparison will once again be restored.

Epistemic functionalism

But okay, okay, I'll admit, not everyone is going to be persuaded by my hardline view that all affective states are functional and contain no non-functional components, but I can sweeten the pot, or rather, remove a lot of the vinegar.

We can weaken functionalism considerably from a claim about how things are to a claim about what it is reasonable to believe (an epistemic claim). If functionalism is the principle that it is a metaphysical truth that no functional differences=no mental differences, epistemic functionalism is the view that it is at least reasonable to assume that there are no mental differences where there are no functional differences unless shown otherwise. Epistemic functionalism is a weaker claim, functionalism implies epistemic functionalism but not vice-versa. By making our premises weaker while still trying to reach the same conclusion (a common strategy in philosophy), we're trying to make an argument that's appealing to a broader circle.

To further explain epistemic functionalism, let's go back to the example of color experiences (even though it's not strictly related to the problem we're considering here). It could be that you see green where I see red and vice versa, but until someone comes up with evidence of that, it's not irrational to think that your green is much like mine and vice versa. connecting it to our topic, perhaps it is possible that all your emotions or all your desires are on a different scale to mine, but epistemic functionalism suggests that we can reasonably assume they are similar in the absence of contradictory evidence.

I'm going to label the rest of this essay as an appendix because it gets more complicated from here on out, and I think that for many people the arguments I have made thus far will go through. Nonetheless, keep reading if you want to learn how we can weaken the assumptions we've made even further.

APPENDIX: TWO EXTRA ARGUMENTS

If even still you don't accept this?: Unbiased estimator functionalism

I find the argument so far persuasive as a solution to the interpersonal utility comparison problem. I'm a functionalist. I think mental states are definable in terms of functional relationships with behavior, and hence are fully psychometrically measurable. Even if I weren't a functionalist, I would find epistemic functionalism or the view that it's reasonable to assume that similar f-states equals similar e-states in the absence of contradictory evidence persuasive.

However, I think we can add another layer of "even if". Even if you find all of the above reasoning unpersuasive, an old argument called the equal ignorance argument, combined with an even weaker form of epistemic functionalism that I call unbiased estimator functionalism, might still go through.

Unbiased estimator functionalism: The equal ignorance principle

Let's suppose that earning another dollar always makes you better off- this means that your utility function is strictly increasing in dollars. Let's further suppose that, despite this, each additional dollar is worth less to you than each previous dollar- this means that your utility function is concave in dollars.

Now let's suppose that a decision-maker knows that every single person in the population has these features- utility which is both strictly increasing and concave in dollars. However, the decisionmaker has no further information on the utility functions of the population- their shape or magnitude. The decision-maker has a pool of money that they want to dole out to the population, how should they divide it?

You may be able to see the answer- which Abba Lerner proved mathematically-intuitively. The decision-maker should split the money equally.

But the equal ignorance theorem suffers from a problem if it is meant to apply to real life. Arguably we are not totally ignorant about the scale of other people's utility functions. For example, arguably we have good reason to think that, on average, rich people like money more than poor people because all other things being equal, the person who likes X more will have more of it. We might also have other reasons to think there are differences. For example, the rich might be more habituated to their wealth- this could create in them a greater need for it or might mean that it is largely wasted on them.

Unbiased estimator functionalism: The claim

By an unbiased estimator here we mean something slightly different from its normal usage. We mean an estimator that, in expectation, is not biased towards higher or lower values for any person or type of person. We still may not trust the answers it gives, we may think it tends to be wildly inaccurate- but not in the form of a known bias.

Let's say you have two friends named Alice and Bob. If you tell either of those friends a story about someone, they will then guess that person's height (it's very irritating). They are both very bad at it and are on average wrong by two feet. However, when Bob makes predictions, he tends to overestimate the height of men, and underestimate the height of women. Alice has no specific tendency towards underestimating or overestimating the heights of men or women. She's just all over the shop. Alice is an unbiased estimator.

Here's where unbiased estimator functionalism comes in. According to unbiased estimator functionalism, a person's functional state can be used as an unbiased estimator of the intensity of their affective states. F-state is an unbiased estimator of e-state.

Glossing quite a bit, the argument from this to interpersonal comparison goes as follows. Let us suppose that we think f-states are measurable and interpersonally comparable, but that e-states are what matters in ethical terms.

If we accept that f-states are unbiased estimators of e-states, even if we don't think they're necessarily very accurate estimators, and that we have no further information about the relative magnitudes of e-states, then, using reasoning like that involved in the equal ignorance argument, we can derive the conclusion that our estimates of e-states based on f-states should guide our ethical behavior. We may be wrong, but our wrongness has no tendency to go in a specific direction and we have no further information so we can't do better than just using f-states as an estimate of e-states.

One final even if

Maybe you've read through all the above and you find yourself radically uncertain about whether or not the arguments I have given "go through" in preserving interpersonal comparisons. I think the argument has been compelling, but I get stuff wrong all the time. Well, I have one final pitch for you, a pitch to commonsense.

Denying interpersonal comparisons is a kind of skepticism. There are all sorts of arguments for all sorts of forms of skepticism in philosophy. For example, some people think that we should be skeptics about induction- or making inferences about the future on the basis of past observations. These people point out that the principle of induction- that it is legitimate to make such inferences- is itself undefended and it's not good enough to say that it must be true because it's worked in the past, because that's circular!

Generally speaking, when we run into a theoretical argument for skepticism in philosophy, we don't allow it to stop actual everyday and scientific inquiry. We assume the philosophers will work it out someday, and even if they don't, we likely keep going all the same.

We have a practically applicable method for making interpersonal comparisons of affective states (psychometrics) that lines up with commonsense ideas about how to measure a variable between instances, and how to validate that measure. Even if you believe that there are philosophical reasons to be skeptical of psychometrics as a strategy for interpersonal comparison, I propose we should keep using it in the interim, much as we keep assuming the future will resemble the past, other people have minds, there is an external world, and so on.

The Paradox of the Crowd

Consider the tension between these statements:

1. The majority opinion in almost every field is more likely to be correct than your own, if your opinion deviates from the majority opinion. This is true even if the group has no more raw data than you—because the aggregate reasoning of the group is likely to be better than yours.
2. If, on the basis of (1), everyone reasoned that they should simply adopt the majority view, the quality of the majority opinion would fall. (This point echoes Grossman and Stiglitz's "On the impossibility of informationally efficient markets")

I take it that it is obvious that the above statements are usually true. (1) may not always hold—for example you might be massively better informed and better at reasoning about a topic than everyone else with an opinion on it. But in the main, individuals are not more likely to be correct than the majority opinion. Even if you think you are an exception because you are better informed and cleverer, which is more likely- that you truly are an exception, or that you are one of the many people deluded into thinking you are an exception?

Admittedly, this is all very context dependent. If you're the only biologist in your society and everyone else has no scientific training and is a creationist, you probably shouldn't be worried about the above paradox, because you'll simply reject (1). The case we have in mind is instead that of an inquirer in a community of relative epistemic equals.

To dramatise the paradox: Galileo said that "In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual." This is wrong, the humble reasoning of a single individual is usually not that good. What is true is that without the humble reasoning of individuals and small groups, the authority of a

thousand would not advance.

One way forward here is to create two sets of propositions. The set of propositions we 'believe' in the sense that we conduct our investigations on the basis of them, and use as the basis of our arguments internally to a community of inquiry and a second set of propositions we 'believe' in the sense that, if you were asked what was all things considered most likely, you would assent to. We populate the first category with propositions that seem to us to be true on the basis of all available evidence except the evidence of the judgments of others. We populate the second category with propositions that seem true on the basis of all available evidence including the judgements of others. Of course belief partitioning in this way may not be psychologically viable for individuals. One alternative would be to give individuals tacit permission to engage in self-deception about the likelihood that they've grasped something the majority hasn't. Then there are hybrid models, where we put some weight on the reasoning of others, but not as much as it probably deserves. This would have the effect of preventing too much ink and lucre being spent on fringe ideas, while still alleviating the paradox of crowd judgment.

Part of orienting ourselves in this landscape is to reflect on our goals. Are we trying to be right, or are we trying to make the group that we are a part of right? Traditional epistemology has assumed that the goal of the agent is, or should be, to have correct beliefs about the world. To this end they seek to form justified beliefs. What if instead we view our goal as trying to expand the knowledge of the group as a whole? This can be quite a liberating way of seeing things. Got some eccentric hobby horse ideas? Excellent! Someone needs to follow those up. Act like you don't know how much of a stab in the dark it is, or even fool yourself into believing they're likely true if it helps. It's all to the good and expands the cognitive reach of the group.

We now enter the domain of Normative Social Epistemology, the study of reasoning for, and as a part of, a group of enquirers, with the aim of supporting that group in its

collective search for truth.

Why I left philosophy

1. Working on intuitions

I started working on intuitions. To see what a philosophical intuition is (or rather, what one type of philosophical intuition is), consider the following:

You might think knowledge is justified and true belief. But suppose I look at my watch and it says the time is 12:37. On this surely reasonable and justified basis I believe that the time is 12:37, and indeed the time is 12:37. However, unbeknownst to me my clock is stopped. It just so happened to stop at 12:37, and by coincidence this happens to be the time now.

Many people have the intuition that in such a case you do not know that the time is 12:37, but you are justified in believing it, your belief is true, and you certainly do believe it. Thus, they argue, having a justified true belief does not guarantee knowledge. If this is true, it overturns what was the almost universally accepted view of what knowledge is for almost two and a half millennia—that knowledge is justified true belief, often shortened to JTB. This sense of wrongness about the idea that the person in the example knows that it is 12:37 is a paradigm case—perhaps the defining example—of a philosophical intuition.

A philosophical intuition is typically (and these are neither necessary nor sufficient conditions!) a sense of rightness or wrongness about the application of a predicate—for example “Knowledge” in a hypothetical case. This sense of rightness or wrongness does not seem to rely on anything external to itself for its own justification, rather it just sort of seems self-evident.

But why are we confident in our intuitions? Why should these sorts of intuitions count as evidence? Do we all share the same intuitions about various philosophical topics? During my undergraduate years, debate had erupted in the literature about what is

sometimes called experimental philosophy. Experimental philosophers run experiments that seem to suggest a diversity of sometimes contradictory intuitions between cultures, social groups, and even within individuals about the same cases. For example, there is some evidence that East-Asian subjects actually regard Gettier type cases as instances of knowledge—i.e., they would grant that the person whose watch says 12:37 knows that it is 12:37, even if this only happens to be the correct time by accident. I wrote a senior thesis arguing against two views on what these experiments meant.

On one hand I argued against people who say that these results debunk the method of hypothetical cases philosophers' use. On the other hand, I argued against philosophers who contend that the intuitions of non-experts about these matters should not be trusted. My argument was informed by an underlying thesis about what intuitions were. I regarded intuitions not as glimpses into a platonic realm, but as constitutive and at least partially stipulative. It was my view that intuitions do not track the truth about philosophical questions, but rather that they are partly responsible for making the truth about philosophical questions.

To see how, let us come back to the Gettier case—although what I say here could apply to many other debates in philosophy. If Janet has an intuition that the Gettier case is not knowledge, this expresses Janet's predisposition to define knowledge in such a way so as to exclude these cases—that she refuses to use the word “knowledge” in a Gettier case partially constitutes what she means by “knowledge”.

If Jiang has a conflicting intuition, Jiang is demonstrating that she is attaching a different concept to the word “Knowledge”. To talk about Janet being right or Jiang being wrong would be nonsense, because their intuitions pick out different concepts. It would be a bit like an argument about whether it was raining on Wednesday where one person means last Wednesday and the other person means this Wednesday. Jiang and Janet could argue about or investigate whose version of the concept was closer to the typical version, and what sort of people use what variants from the big family of

Knowledge concepts—and to do so they might find experimental philosophy useful. They might even debate which concept of knowledge was more useful but debating which is correct is meaningless. Each of the two concepts describes or picks out a different property. The role of experimental philosophy then is to show us the variations in the concepts people are deploying—it is a project of philosophical lexical semantics.

A lot of people seemed worried that this approach would lead to anything goes relativism. I don't see the argument. If Jiang has X intuitions about the concept of knowledge, and Janet has Y intuitions about the concept of knowledge, then they are using slightly different words. There will still be a fact of the matter about whether someone's belief is knowledge as Jiang means it, and there will be a matter of fact about whether someone's belief is knowledge as Janet means it. Once propositions are properly disambiguated, there's no spooky 'the world is just a point of view' relativism going on.

There's a common cousin of the view I've outlined that I'd best explain so as to distinguish it from my own. It goes like this. "Philosophers can't study knowledge itself using intuitions, they can only study people's ideas of knowledge. There is this further thing which is Knowledge-itself and either philosophers can't study it, or they need to use some method other than intuitions about cases to do so". I can see why people would confuse this view with the view I've outlined, but I think they're quite distinct. Here's why: If there is a rich enough infinitude of properties in the world, then for any meaningful concept there will be a property corresponding to that concept. As a result, if we come to fully map out someone's intuitions which define a concept C, then we will also find out exactly which property C picks out. If we fully map out a folk concept, then, at least prima-facie, we also fully understand its corresponding property in the world. There's no need to say things like 'we understand what this group of people mean by knowledge, but not what knowledge itself is—there's some further fact about that.' What the property of knowledge is, is given fully by what the concept of knowledge is.

The character of the property of 'being knowledge' just falls right out of an analysis of the concept of knowledge.

The view I've outlined has many advantages. It avoids tricky epistemological puzzles, for example: why should we think intuitions tell us something about an intangible metaphysical world external to our own minds? If intuitions really do systematically vary between cultures, it avoids the awkwardness of having to explain how one group came to be right and the other group wrong. Also, when you think about it, the idea of one property out there in the world being Knowledge with a capital K is kind of silly. There are of course also arguments against it—mostly in a bundle of ideas related to reference magnetism, direct reference and the causal theory of reference which I won't get into here. There are also concerns built on Quine inspired semantic eliminativism.

Disclaimer for experts: So it doesn't seem like I'm denying certain facts let us clearly acknowledge that nothing in this view says that some linguistic frameworks aren't better equipped to describe the world, or carve it at its joints, or simply be more useful, than other frameworks. Let us also acknowledge that what is 'useful' will be context dependent. Let us acknowledge even further that nothing in our view denies that there may be a degree of reference magnetism towards the 'joints' of nature, it only requires that it not be strong enough to outweigh the possibility of alternative or parallel concepts for important philosophical topics like 'knowledge', 'mind' and 'personhood'.

As far as I know there is no name in the literature for the view about what intuitions are, and what studying concepts really does, that I have described. Despite that, when you talk to philosophers it becomes clear that the view I've described is extremely common. Indeed in some groups, something like it seems to be the majority view, which makes the absence of an explicit name for the view all the more mysterious.

The view described here some has similarities with ordinary language philosophy, though few ordinary language philosophers stated it so baldly, and it's hard to tell since,

following Wittgenstein, so many ordinary language philosophers seem to have made it a virtue to not be clear on exactly what it was they were doing. It has some definite similarities with what is sometimes called the Canberra Plan. Alvin Goldman outlines a vaguely similar view, although from memory his view is a bit more like the common cousin I outlined a few paragraphs above.

Nonetheless the view is rarely argued for in explicit terms. To revise after so many years I reread the Stanford Encyclopedia of philosophy article on intuitions before writing this—nothing like the view that intuitions make, rather than track, philosophical truths is described, despite exhaustive discussion of the minutiae of various other views about, and aspects of, intuitions.

As a step towards overcoming this marginalisation, let us call the view constitutivism about intuitions. Formally stated then, constitutivism is the view that intuitions are not merely truth-trackers, but truth-makers regarding certain philosophical claims. The reason for this truth-maker status is that a person's intuitions help constitute what they mean by a term in their idiolect, and philosophical questions on the constitutivist view often turn on semantics. Further, from the content of concepts we know trivially the nature of the properties they pick out, for any meaningful, complete and non-self contradictory concept has a property which corresponds to it.

Succinctly—intuitions are part of concepts and different intuitions about appropriate usage give different concepts, concepts give us properties, thus once we've finished conceptual analysis, we've finished. Variation in intuitions, whether between subjects, or even within a single individual, just indicates a plurality of concepts.

2. Collapse

But what does this have to do with me leaving philosophy?

I finished up my senior thesis which briefly touched on these themes, and then began a

doctoral thesis. In the part of the world I was doing my doctorate, you immediately begin researching and ideally writing your thesis almost as soon as you enrol, unlike the American system. Like any prototypical twenty-two year old philosopher I was wildly over confident in the value of what I had to say, and ready to change the world. That's when I encountered the haze. A lack of concrete research on the topic of my thesis

I could tap into, of an accessible bedrock of literature which I could build a thesis on. There were many papers on metaphilosophy tangentially related of course, but everyone seemed to be coming at it from different angles, groups of people were having conversations that slid entirely past each other. There was no obvious way for me to slip into the party with grace. It seemed to me that there were a great many people who thought they were talking about the same things, but really were talking past each other. There weren't even always names for the various constellations of positions people took. I was lost.

There are, I think, a number of causes for the haze. Two of which lay with me, one of which I could be rightly blamed for, and the other I couldn't. The one I could be rightly blamed for was that I was nowhere near as fucking smart as I thought I was. The one I can't be blamed for was that I was severely mentally ill. However there were also external cases for the haze.

First of all, many philosophers just don't care that much about metaphilosophy. When I talked about my thesis with faculty members and fellow students I generally got one of four responses:

“This is obviously true.”

“This obviously not true.”

“This is too abstract to be interesting.”

“That’s really interesting, I’d never thought about it like that.”

I realised with astonishment that many—though not all—philosophers had the same orientation towards metaphilosophy that many scientists have towards philosophy. Just like a room of scientists asked about science would tend to throw up sentiments like:

“That’s all good and well, but what’s the practical point?”

“I have strong opinions on this topic but haven’t thought about it that much.”

“Ooh that’s really interesting but it’s a little abstruse for my taste.”

So too did philosophers. The other cause of the haze was that even though what I was saying corresponded to what a lot of philosophers believed no one had thought to name it. And for some reason, even though many philosophers are instinctively drawn to the idea, people who work on topics like ‘what are intuitions’, don’t seem to much like the Idea.

I began to wonder if maybe it wasn’t because there’s vaguely self-effacing about the whole thing. If you sort of suspect that parts of philosophy (not all, mind) are continuous with semantics—“semantics” being a byword for irrelevancy in our society—maybe you sort of want to shut up about it. Thus you’ll only write on relevant topics in meta-philosophy if you think you’ve got something to say which carves philosophy into a more imposing figure.

I guess I never saw it that way, because I think there’s a kind of glory to be had in studying such important concepts and words as knowledge, personhood, belief, desire and so on. Some questions about the semantics of some words and the structure of some

concepts go to the very heart of the way humans understand their world, and are humans not noble in reason and infinite in faculty?

Overwhelmed by a conversation that was at once fragmentary, vast, and hard to find, in which people who agreed with me never seemed to speak up, which many philosophers seemed to think was a useless conversation, and sapped of energy by my health problems, I sank out of academia.

How to do things to words: mapping a post-analytic philosophy of concepts and intuitions

Introduction

I wrote my honours thesis on experimental philosophy, almost a decade ago. I then went on unsuccessfully to attempt a PhD. My feeling at the time was a feeling common to many philosophy undergraduates, but usually eventually beaten out of them. It seemed to me that many debates in philosophy were really, at heart, semantic or merely verbal debates.

A number of developments in the philosophical literature- from experimental philosophy, to the development of the idea of conceptual engineering (e.g. Chalmer's recent paper on conceptual engineering, which this post owes a great debt to and Haslanger's paper on race and gender) have led me back into this topic. I wanted to lay out, in simple English, a few thoughts I've been working on for years about the objectives of a post-analytic philosophy. Post analytic in the sense that analyzing concepts would not be a central objective, at least where analysis is conceived of in the normal way.

I lay out a number of ideas. Some of these ideas are mine, others are not. A lot of it it is stuff I've thought of independently, and later found others also thought of. The lack of citations does not mean I'm claiming credit, it just means that I'm lazy, and this isn't an academic paper.

A peculiar game

A big part of what we call analytic philosophy is the following game. I try to give necessary and sufficient conditions for a thing, and you find counterexamples which

either A) exemplify the thing, but do not meet the conditions or B) Do not exemplify the thing but do meet the conditions.

Famously Plato proposed that man was a featherless biped and Diogenes responded by shaving a chicken

Let's suppose we are debating the definition of "life". I propose a definition- a set of necessary and sufficient conditions- life is any process that reproduces itself. You

respond by reminding me that fire reproduces itself, yet is certainly not alive. Perhaps I parry by adding to my definition that the putatively living thing must also achieve homeostasis. In this way the game continues.

Empirical evidence suggests that these games are interminable. No one ever wins. Even definitions that seem solid don't hold up. Knowledge as true, justified belief was one of the few "wins" of the methods, but as we'll see below, that was overturned by Gettier in the 20th century.

What we're going to do

In this essay, I want to lay out what I take to be three theses or phenomena that taken jointly undermine the tenability of this game as a pillar of philosophy, viz:

1. The stipulative theory of intuition
2. The family resemblance approach to concepts and
3. Diversity of intuitions.

Then I want to talk about some different activities which are an alternative to this game of trying to find necessary and sufficient conditions.

The stipulative theory of intuition

We'll use knowledge as an example. One (formerly) popular definition of knowledge is that it is justified true belief. That is to say, for something to count as knowledge you must:

A) Believe it B) Have an adequate reason to believe it (justification) C) And it must be True

This seemingly simple list of requirements held up as a definition of knowledge for almost two and a half thousand years. However, there is a counter-example, owing to Edmund Gettier. Suppose Bob looks at his watch and it says 6:30. It is in fact 6:30. Bob's belief is true (because it's 6:30) and justified (because a watch provides adequate justification for a belief about the time). However unbeknownst to Bob, his watch is stopped- it stopped several days ago. It just so happens to be stopped on the right time by chance. As a result of this counterexample, relatively fewish practicing philosophers today believe that to in order to count as knowledge, it is enough that a belief be true and justified. Additional conditions are thought to be necessary, although the nature of these conditions- and even whether there is a compact, simple list of conditions, is a matter of ongoing debate.

There are, I think, at least two important stories about what this Gettier intuition, and by extension, many similar intuitions about a huge range of different topics, are doing:

The theoretical account: On this account, our intuition that the Gettier case isn't a case of knowledge is like a hunch that a certain claim about the world is false. That hunch can be wrong.. Just like a belief that pyrite is a type of gold is wrong, so it could be wrong that Gettier cases aren't a type of knowledge. This hunch is thought to count as a type of evidence about the real nature of knowledge, for reasons that have never been entirely clear to me.

The stipulative account: On this account, our intuition that the Gettier case isn't knowledge isn't a theory about the world. Instead it's part of what fixes the definition of knowledge. Let's say that I define a Xuzazis as an object which is at least 50% green and weighs at least a kilo. I couldn't then find out that a Xuzazis can weigh less than a kilo. Similarly my "intuition" that the Gettier cases aren't knowledge is really me stipulating that any property whose extension includes the Gettier cases is thereby not the property of being knowledge.

It has long seemed to me that philosophers are insufficiently clear about which of these two accounts they think is true. I think this an important metaphilosophical dispute which deserves a lot more attention. Of course there are debates which relate very directly to this-debates over the causal theory of reference, over the Canberra plan etc., but it still seems to me that this is such a central question it should be discussed more, and more explicitly.

What I call the stipulative account has many advantages over the theoretical account. For one thing, it makes the epistemology of intuitions much less mysterious. The theoretical account has difficulty explaining how intuitions can be a guide to truth without resorting to mysticism. On the stipulative account there is no mystery. To the extent that intuitions are even beliefs at all on this account, they are beliefs that are their own truth makers.

For the sake of argument I accept the stipulative account as true, at least of very many important philosophical debates, for the rest of the essay. I acknowledge the possibility that the stipulative account is not true of everything. Philosophers have long noticed that certain words seem attached fast to certain natural kinds- like the word "gold" and the chemical AU 79, but I think there is an important class of philosophical debates deploying linguistic intuitions which the stipulative account is adequate for.

Family resemblance

What if concepts don't come with compact lists of necessary and sufficient conditions?

What if what we're looking for is more like a family resemblance? No one feature can rule you in or out.

One way this might work is through prototypes- the so called prototype theory of concepts. Maybe when someone says that X is a "bird" for example and asks whether we agree, what we do is compare X to a prototype bird- perhaps a hawk or a robin, checking for properties. The eagle passes fairly easily. The penguin leads to greater hesitation. The emu, even more so. There is considerable empirical evidence in favour of the prototype theory- although far from decisive. Note though that the semantics of words and the extension/intension of concepts can work like family resemblance even if the prototype theory of how we process concepts isn't true.

Were inclusion in a concept more like family resemblance than a set of necessary & sufficient conditions, the classical philosophical approach to analysis would no longer be possible. Interestingly though, there would be an alternative, call it quasi-analysis. Quasi-analysis is the practice (purely hypothetical as far as I can tell) of laying out those features that tend to make something more X like- but not as necessary or sufficient conditions. So, for birds, a quasi analysis might include: Wings, feathers, beak, lays eggs, can fly, builds nests, sings, squawks...

For knowledge it might include:

Is a belief, is true, is justified, is held very confidently, is justified by beliefs that are themselves knowledge, is widely agreed upon by those considered qualified to assess it, is certain, has been formed by a reliable mechanism, does not contradict other things the agent believes...

And so on...

Diversity between and within people

Experimental work on intuitions has revealed that there are systematic differences

between intuitions about philosophical questions:

- A) Between social groups (genders, races, classes etc.)
- B) Between the same individual within different contexts (emotional state, disposition to blame etc.)
- C) Idiopathically (two individuals in the same groups might have different intuitions just because- or maybe due to personality facets)

For example, coming back to knowledge again, experimental work on intuitions (sometimes called “experimental philosophy”) has revealed that there may be both cultural and contextual influences on whether or not people consider the Gettier cases to be cases of knowledge. I believe also that there are variations along the lines of personality, and also of situation. Result A makes it impossible to analyse “the” concept of anything, since there is not just one singular concept. Result B makes it unlikely that we can even hone in on a specific demographic group, and study their concept of X, because it is quite likely they have more than one, varying between situations. Result C puts a bow on top. The joint effect of diversity between and within people, the family resemblance approach to concepts and the stipulative theory of intuitions. The joint effect of the three propositions I outlined is to make the game of hunting for necessary and sufficient conditions quite futile, although I don’t think any of these points does it alone.

If just the stipulative approach to intuitions were true, and the other two propositions were false, we could keep hunting for necessary and sufficient conditions as an exercise in understanding the concepts in people’s heads.

If just the family resemblance approach to concepts were true, it would be unlikely we’d find necessary and sufficient conditions that just so happened to meet all cases, but we might still learn some interesting things by playing the game, even if it were unwinnable. We might even say that, even if concepts work on family resemblances, it

could turn out that some of them have relatively compact, workable necessary and sufficient conditions “by chance”.

If just the diversity thesis were true, we might simply say that we had a lot more concepts to track- the different concepts of each culture, situation etc. We might even say -very brashly given the history of such things- that some cultures or people in some situations, had more correct approaches to certain concepts than others. But I think the overall effect of these three propositions combined is to make playing the necessary and sufficient conditions game if not useless, at least of limited utility. Once we accept these premises , what linguistic and quasi-linguistic philosophical tasks remain? In what follows I lay out a number of linguistics and quasi-linguistics tasks that remain once we accept these postulates. The tasks I lay out are, at least broadly, philosophical tasks. My favorite task is at the end, so keep reading.

Concept creation (CC)

In concept creation as the name suggests we create a new concept. Probably ideally this is to go with a new word, but it might go with an old word as well, as a new meaning of that word. An example is my word Yvne, the inverse of envy. Yvne is cruel satisfaction that others are deprived of something you have, or have less of it than you do. Definition creation does not seem like an especially philosophical task on the surface, although on second thoughts finding blind-spots in our web of concepts and filling them maybe is a very philosophical thing to do.

Another great example of a philosopher creating a concept to direct our attention to something missed in ordinary thought- Tamar Gendler’s concept of Alief’s. To see what an Alief is, imagine standing on very thick and sturdy a glass floor over a deep ravine, going down hundreds of meters. You likely believe that you are safe, but you alief that you are not. Similarly if you are eating chocolate fudge shaped like faeces you likely believe that this is hygienic, but do not alief it. This is an example of philosophically

interesting and provocative concept creation. [Chalmers seems to have thought of this as an example independently- too late to edit it out now]

Conceptual zoology(CZ)

There are a lot of already existing concepts of philosophical interest, waiting to be discovered by philosophers. Sometimes these exist as alternative uses of philosophically loaded terms- and thus have remained hidden from philosophers, who have seen them as deviant usages rather than appreciating them on their own terms. There is a lot of work to be done discovering, classifying and understanding the role of such alternative concepts. Consider, for example, what we might call the sociological concept of knowledge, commonplace among those who study the sociology and history of “knowledge” of various sorts. Here knowledge means something like socially sanctioned belief. Or at least this seems to me to be the definition at play. This concept of “knowledge” itself has various subtleties, and is worth the trouble to try to understand- and not just treat as a postmodern knockoff of the real thing.

We might also suspect that there is a scientific concept of knowledge. On the scientific concept of knowledge, a proposition can be “known” even if it is not really “believed” as such, or even true- it counts as knowledge just so long as we are justified in provisionally

accepting it. We say that we know stuff to be true on basis of it following from relativity theory, even though it is quite likely that in a better, future science relativity will turn out to have been only a approximation. The proposition is thus unlikely to be true, not really believed, and only in a sense justified, yet it still would not be too strange to call it knowledge!

Conceptual redefinition(CR)

In conceptual redefinition I redefine a term for some purpose. The degree of redefinition can vary. I might try to capture what I regard as really meaningful about the term, or I might make something very different.

For example, “by knowledge, I mean justified true belief- inclusive of the Gettier cases” would be a conceptual redefinition of knowledge. A more radical reconstruction would be “by knowledge, I mean any correct belief, even without justification”.

Here are some of the use cases for conceptual redefinition:

Social recognition: When gay marriage was still a goal, I would sometimes argue with conservatives who said that the common-sense definition of “marriage” was that it was between a man and a woman. Obviously I didn’t accept this claim, but one of my favourite responses was that, were that true, we should change the definition for the sake of recognising an important group of people and their relationships.

Analysis: In the past I’ve suggested altering the term envy to include both what is currently called envy and what I call yvne. On such a redefinition, envy would be “a preference that others do poorly relative to yourself regardless of whether those others are currently above or below you”. Such a concept, I think, would be useful for seeing the world as it currently is. The current concept of envy is biased in that it focuses blame on those who are at the bottom of the social heap. In that regard it is ideological it represents the fear the powerful have towards their lessers, and conceals the truth that the rich can often desire the failure of the poor as much as the poor desire the failure of the rich. This is an example of championing a conceptual modification for purposes of clarifying analysis. In this example the analysis is social, but it could just as easily relate to the natural sciences.

Removing ambiguity: We can imagine a philosopher who, with a certain purpose in mind, declared that, henceforth by knowledge he would mean true, justified belief, even

inclusive of the Gettier cases.

Normatively guided redefinition (NoGR)

This is a special case of conceptual redefinition where we try to make a definition correspond to a normatively significant category. Suppose I were trying to come up with a definition of “torture” for example, I might be focused primarily on a cluster of behaviours that are generally bad for the same reason. Maybe, for example, ordinary people don’t use the word torture in such a way as to capture imprisonment, but I think imprisonment is in all morally relevant respects like paradigm cases of torture.

Therefore I redefine torture to include imprisonment, on the grounds that this doesn’t create distinctions without a moral difference.

The normativity doesn’t have to be moral. Maybe I think that, although the Gettier case shows that justified true belief is not always knowledge. Nonetheless, I think justified true belief is always as epistemically praiseworthy as knowledge. I, therefore, propose that we should, either in a specific context or maybe even generally, redefine justified true belief as knowledge, because it matters and is valuable in the same ways that knowledge matters and is valuable.

Another potential example of normatively guided redefinition is the concept of survival as in that person survived that event. Or to put it another way, the temporal boundary conditions of the concept of personhood. For example, philosophers have long argued over whether one would count as surviving if one’s body were disintegrated and reconstructed through a teletransporter. Increasingly an increasingly common view, argued by authors like Parfit (c.f. Miller for a similar position) is that this is the wrong question. Our intuitions about whether we survive this or that are hopelessly confused and unlikely to be turned into a single coherent narrative. Instead we should ask what do we care about? Mental continuity seems to me to be what I care about, regardless of whether you call this survival. Perhaps you are different though.

Natural kind hunting (NaKH)

This is a kind of extra-linguistic project that ties into the linguistic projects we're talking

about here. According to the Stanford Encyclopedia of Philosophy:

“To say that a kind is natural is to say that it corresponds to a grouping that reflects the structure of the natural world rather than the interests and actions of human beings[...] Putative examples of kinds may be found in all scientific disciplines. Chemistry provides what are taken by many to be the paradigm examples of kinds, the chemical elements...” In natural kind hunting, we look for natural kinds which share a similar extension to philosophically loaded words in our language. In some cases it might even be possible to find natural kinds which correspond exactly to our words. Historically this has often been done by people who think that natural kinds act like magnets for our words- but it doesn't have to be.

For example, I could uncover that there's a particular kind of brain state that corresponds to many, but not all, uses of our concept of belief. This would be a philosophically interesting discovery. We need not believe that it is revealing or changing anything about the definition of belief. Whether it does or it doesn't it is still, I think, an interesting scientific and philosophical task that relates to meaning.

To sum: NaKH might or might not be associated with a proposal to create a new concept which more precisely matches the natural kind, or with a proposal to reform an existing concept so that it matches the natural kind- but then again, it might not.

Natural kind hunting is interesting in and of itself, and for many possible natural kinds (like those related to folk psychology- belief & desire), philosophers will have a lot to say in the hunt.

Philosophical lexicography (PAL)

We come to my favourite kind of project, which I call Philosophical Lexicography. Philosophical lexicography is a research program, continuing on from experimental philosophy, which aims to:

- A) Map the usage and the variations in usage of philosophically important terms between groups of people and between the different contexts individuals find themselves in.
- B) Understand these similarities and differences in terms of cognitive needs -universal and specific-, material circumstances -universal and specific-, personality factors, cultural factors, the history of ideas, the evolutionary history of our species, etc.

I have no doubt that this project of philosophical lexicography will be misrepresented as a relativist project- a kind of postmodernism in scientific garb. This isn't fair though. If Bob has a different concept of knowledge to Alice, for any given belief, B, there will be a fact of the matter about whether B is knowledge in Alice's sense, and a fact of the matter about whether B is knowledge in Bob's sense. There's no real relativism going on here. Different people mean different things by the same words, but we can hold the meaning fixed if we like, and there's only one reality that the words and meanings are being matched against.

Others will suggest that this project is all good and well, but that there remains a further fact about what knowledge really is, aside from our conceptions of it. I suppose there are ways this could turn out to be true but I see little reason to believe it, anymore than I find reason to believe there might be a xuzazis that weighs more than a kilo because there is a one true concept of xuzazis outside our heads.

Afterword for suspicious philosophers

I love conceptual analysis. I love playing with the intricacies of words. My own education and sympathies lie with the Canberra Plan. My real intention here is not so much to kill conceptual analysis, as to find a suitable afterlife. I've long disliked both the brash Quinean perspective of Epistemology Naturalised and the brash approach of trying to get intuition out of the picture by turning everything into a natural kind and combining it with externalist semantics. The kind of project I've outlined here leaves room for a paradise of Gedankenexperiment and counter-Gedankenexperiment, while not pretending that we're ever going to find necessary and sufficient conditions for anything unless of course, we declare them by fiat.

Against Libertarian Criticisms of Redistribution

Pt 1: Non Aggression tells us nothing about the morality of redistribution

According to the non-aggression principle one should never interfere with the person or legitimate property of another without their permission, unless they have initiated aggression against one first. The non-aggression principle is sometimes taken to be a master argument for libertarian views against the redistribution of money or property—e.g., left wing proposals to redistribute money from the rich to the poor. I won't argue either for or against the principle of nonaggression, as there are far more pressing ethical issues. Instead I'll be contending that the non-aggression principle tells us nothing, at least directly, about the topic of redistribution.

In the definition of the non-aggression principle I insisted that the non-aggression principle applies to legitimate property. I'm not trying to smuggle anything especially controversial in here. By insisting on the term legitimate I'm simply insisting that you actually have to rightfully own the thing in question, it's not enough to simply proclaim that one owns it. A moment's reflection will show that this stipulation is necessary, if one owned everything one proclaimed one owned then many things would have multiple inconsistent ownership claims.

Consider the case of Bob. Bob passionately claims that he owns the Atlantic ocean, he actually seems to believe this, and insists that no one should cross the Atlantic without his permission. When asked to justify this, he responds by saying that crossing his ocean without his permission is aggression, and everyone should accept an ethical norm against aggression. When confronted with this argument, there is no need to say anything for or against the non-aggression principle, one simply has to say that the Atlantic Ocean is not actually Bob's, therefore no aggression against Bob has occurred. This is where the champion of the non-aggression principle as a basis for

libertarianism hits a problem. The supporter of redistributive taxation typically does not accept that the goods and monies to be redistributed are, in fact, the legitimate property of those they are being taken from. They hold, on the basis of a differing theory of distributive justice than that held by the libertarian, that they are the rightful property of someone else.

The libertarian will respond by insisting that, yes, the prior owner is the legitimate owner of the goods or monies in question, but notice that the argument has now strayed beyond the issue of non-aggression into a debate about who owns what. Our point is simple then, non-aggression tells us nothing about redistribution unless we assume that redistribution is a process of removing something from its rightful owner and giving it to someone else but this is part of what is under dispute in debates about distributive justice. The debate is really about who is the rightful owner of what, and unless one can win this debate, one might as well be Bob insisting that he owns the Atlantic. Just as there is no aggression against Bob implicit in sailing across the Atlantic ocean and 'breaching' his sovereignty over that ocean, so perhaps there is no aggression in 'taking' money off Bezos to pay for redistribution, if the recipients of that redistribution are already the rightful owners of that money.

Put simply, taking your stuff is not aggression unless it actually does rightfully belong to you, and the whole project of the advocate for redistribution is to try and prove that, in some cases, it doesn't.

In fact if the supporter of redistribution is correct about who rightfully owns what, then in the non-aggression principle would imply that action resisting redistribution is impermissible, as it would be a form of aggression.

Now of course the libertarian has responses to the advocate for redistribution. They can critique the arguments in favour of redistribution and propound their own theories of who owns what that do not allow much of a role for redistribution, for example, as

Nozick does in Anarchy State and Utopia. However such arguments are not primarily appeals to non-aggression, rather they are theories of who owns what. Non-aggression simply doesn't cut at the difference between the libertarian and the redistributionist.

Pt II: History and Property Rights

In the previous part I outlined why I don't think non-aggression is sufficient grounds to prove that redistribution is bad. I said that what the libertarian really needs is a theory of entitlement to property that provides a defence against redistribution. Of course many libertarians have aimed to produce such a theory. Here are two desiderata for a theory of property rights that can act as the basis of a case against redistribution:

D1. It must justify the existing distribution of property.

D2. It must do so without appealing to the state or any collective body for legitimation.

Let's unpack desideratum one (D1) a little. It is not enough to establish that people have property rights or that property rights are important to fulfill D1. One can imagine proving that there are lots of property rights, but that they do not correspond to the existing socially recognized distribution. Such a conception of property rights would not be a defence against redistribution, instead it would require redistribution. Instead the opponent of redistribution must simultaneously prove that property rights exist, and that they are similar to the existing set of socially and legally recognized property rights.

Desideratum two (D2) is equally important. A proof that property rights exist because the state makes it so would presumably leave the state in a position to change the distribution of these property rights, thus it would be no defence against redistribution. The most common libertarian approaches which aim to meet these constraints are historical theories of distributive justice. These theories typically hold that you are entitled to something if you justly acquired it from nature, or if you acquired it consensually from someone who did acquire it justly from nature, or if you consensually acquired it from someone who acquired it consensually from someone who justly

acquired it from nature, and so on.

I'm sceptical of the claim that any historical theory of distributive justice will ever meet both desiderata because existing regimes of property rights have been arrived at through morally contorted historical processes that libertarians do not accept the validity of.

There is no just chain of transmission for the computer I am typing this on. It was built using raw materials from land that was many times stolen and re-stolen. The company that made it was funded using government subsidies that libertarians object to. Every good was made using numerous other goods, and all of those goods made with many other goods in turn, and you don't have to go back far in the history of anything to find numerous interferences with what libertarians would regard as just.

These problems ripple through the whole. If I purchase something from you that you don't really own, I don't really own it either, and if I then sell that thing in turn, I don't really own the money I gain from selling it and so on. Call this the problem of spreading contamination.

At this point most libertarians I've spoken to have responded with something along the lines of the following:

'look, it's the best we've got—sure actually existing society is riddled with theft and misappropriation, and this can be found in the history of practically any consumer goods, land or capital that is owned by anyone, we have to go with what we've got because the alternative is even more theft and misappropriation.'

The problem here is that by definition it's not actually theft unless you own it, and on the historical theory of distributive justice you don't own it because it wasn't acquired justly. You might say "so you're alleging that taking anything from anyone isn't

theft—that’s absurd.” My reply is that I’m saying no such thing, what I am saying is that, taking the historical theory of distributive justice seriously, this is what it entails. This is not an argument that you can take anything you like from anyone; it’s an argument that we need a better theory than the historical theory of distributive justice to explain why you can’t.

Appendix A: The contention that objecting to capitalism is objecting to private arrangements purely between private individuals

Sometimes Libertarians argue that capitalism is just a series of private arrangements between consenting adults, so there is no grounds to dissent from it. The thing is, most actual exchanges under capitalism involve claims to capital goods and land that society might well contest the ownership of.

Let’s suppose I found a television by the side of the road, dumped there by some thieves, and was about to sell it. Whereupon you discovered me, and explained that, actually, you have a better claim to the television, and so the transaction can’t go ahead.

If I then reply “but this is a contract between consenting adults!” this would be wholly irrelevant because our agreement involves infringing the rights of others. The libertarian begs the question by assuming that the consensual transactions don’t involve trading in things which someone else (say, the state) has a better claim to. This is not to say that the state actually does have a better claim—to work this out we will have to consider difficult questions of moral philosophy and economics, but merely to say that the bare fact that a transaction is voluntary between two consenting adults does not in itself establish its legitimacy.

Indeed the “voluntary” arrangement might even infringe the rights of a party to the transaction. Suppose that you are the buyer, prepared to buy back your own television at considerable expense. In some sense you are a voluntary participant to that transaction, in another sense you aren’t really a voluntary participant at all—you have a right to get

the TV back either for free, or at only the cost of a finder's fee, presumably less than the present value of the television.

Appendix B: The tyrannical king as a benchmark

Suppose that a merchant went before a king and said "My lord, your taxes doth oppress me fiercely." To which the king replied "Begone vagabond! It is only by my clemency that I do not charge you far more! For this whole land is mine. Truly what you pay is not taxes, but a fee I impose on people who dwell on my property. What you think of as "your" land is really mine, and the land taxes you pay are but rent! It is nothing less than the natural laws of liberty which permit me to tax you."

From a deontic libertarian point of view, the only option here is to challenge the king's claims to own everything. I want to suggest that for many purposes we can use the king as a benchmark for assessing the plausibility of Libertarian defences of the status quo. For example, if the merchant sayeth to the king—"But my lord—thou hast stolen all that thou possess, or inherited it from those who have!" and the king responds—"Certainly it is so—but further theft at this point to 'rectify' the situation would just perpetuate the cycle of stealing—best leave things as they are!" This argument would be transparently absurd, yet I've heard libertarians make exactly this argument for why we should not engage in mass redistribution to correct past wrongs, (and things libertarians regard as wrongs, such as government subsidies).

Appendix C: There are no golden strings, just institutions

Although this doesn't prove anything, I think it's useful to take a breath and clear our mind when we think about property. A lot of people imagine property as somehow metaphysically tied to a specific owner by intangible golden threads, and it's worthwhile to remind ourselves that this is not so.

Never forget that ultimately there are just objects. Tables, chairs, parts of land, and

people, which are a special kind of object. What is property then? Property is a kind of social arrangement giving certain people certain bundles of permissions regarding certain objects, and denying those permissions to everyone else. In the final analysis then, like all permissions and refusals, property is a collection of threats of social sanction, including violence.

It seems deeply unlikely to me that we will ever be free of property understood in this way, or that this is even desirable. Even a communist state wouldn't want people trespassing in the nuclear power reactor without the right expertise—and what is the right to collectively exclude all people who lack special permission from a site but a kind of collective property?

Essential though it may be, re-framing property as the threat of sanction and violence, and not some metaphysical linkage, brings it into a new perspective. From this standpoint there is nothing especially 'non coercive' about, say, anarcho-capitalism, unless you take it as given that the claims it makes about who is entitled to what are ethically just.

Through-going subjective Bayesianism as a solution to the problem of skepticism

A common argument for the existence of God is that there is something rather than nothing. There are many good replies to this argument, but one of the more sophisticated challenges our sense that ‘nothing’ is the ordinary state of things and ‘something’ is an exception that needs explanation. There is an enormous, uncountably infinite plurality of ways the world could be. The ‘nothing’ world is simply one of these ways, and no more inherently ‘natural’ than the others. Why must the initial state of the universe be nothing unless we can give a reason otherwise?

We might talk then of the ‘nothingness is natural’ fallacy in relation to the initial conditions of universes. Skepticism seems to me to commit a variant of this fallacy in relation to epistemology. The set of permissible beliefs is assumed to start empty or near-empty, and then we have to fill it. A very low probability judgment for all statements is assumed to be the real correct default. Any exception is in need of justification.

There is an alternative view of epistemology that does away with skepticism. I see it as, broadly speaking, a development of G.E. Moore’s argument against skepticism to a whole epistemology by way of subjective Bayesianism, though this might simply be reading my own ideas into Moore’s work. It is a development of subjective Bayesianism into a full epistemological standpoint, a kind of alternative to coherentism and foundationalism, or a dissolution of the questions they attempt to dissolve, depending on how you look at it.

The view goes like this. There are no rationality requirements beyond consistency on the initial assignment of priors. Rationality only governs the updating or changing of our beliefs and does not impose any constraints on our initial beliefs beyond consistency requirements of the kind given by the Dutch book argument.

The real province of normative epistemology is purely the assessment of changes in probability assignments. The demand for some ultimate justification of the initial judgements is an illusory over-generalisation from the practice of justifying changes in levels of belief. Our priors start out non-sceptical, and so our beliefs remain throughout a lifetime of updating—there is no higher ledge to stand on from which to critique this. On the subjective understanding of probability, we do, after all, have to have some priors, and there is no known principled way of assigning them. Thus we arrive at some rather simple anti-sceptical proofs. For example, mirroring Moore's proof of an external world, we simply observe that our prior belief in an external world is rather high, and nothing has decreased it.

So why does the problem of scepticism appear to be a real problem? Over generalization. If I tell you that Susie is a lecturer in mathematics, you might ask how I came to know that, because you judge that it is very unlikely that it's simply built in to my priors that Susie is a lecturer in mathematics. This is true of basically all statements worth talking about outside a philosophical context, thus we get into the habit of thinking in terms of justifying our beliefs. When I then tell you that I believe there is an external world, and you ask how I know this, and I cannot give an adequate reply, it appears something is wrong. Really though this is a sort of illusion, all we are ever doing is justifying why we have decreased or increased our credence. Since your a priori belief that Susie is a mathematician is presumably low you need to explain why it has risen but the same is not true of belief in an external world, or causation- both of which start with high priors.

I'm not particularly worried about confirmation bias

I. What is our epistemic goal?

If I were worried about being right about everything, no doubt I would be having a rough time of it with confirmation bias. I would start to think about something, form an opinion, become petrified this was merely me seeing what I wanted to see and explore the opposite opinion, then become petrified by the possibility that this was also me just seeing what I wanted to see as well, but from the opposite point of view, a kind of grim truth to wear like a hairshirt, to convince myself of my epistemic virtues.

I'm not in this spiral, largely because I'm not aiming to be right. That is to say my primary aim in inquiry (outside of day-to-day life) is not to have true, justified beliefs. I'll justify this extraordinary claim in a moment, but for the sake of a logical ordering, let me start by saying what I am trying to be. I am trying to be interesting and honest (and you will observe these are at least potentially contradictory). By interesting, I mean bringing to the table ideas that are novel, or, more plausibly, are at least novel for much of my audience. By honest, I mean not misrepresenting or deceiving about the strength of the evidence for these ideas.

You'll notice that both these virtues are other-directed. That is, I have described my approach to thinking in terms of interacting with others. So now we come to why I'm not particularly worried about confirmation bias. My goal isn't to be right individually. My goal is to make my little contribution to the world as a whole getting it right.

Given that overarching goal, bringing novel thoughts forward, while seeking not to exaggerate my case and distort "the discourse" is the best I can do.

In a way, the fact that I am prone to confirmation bias might even be a good thing from the point of view of the world getting it right. It allows me to range deeply in epistemic

tunnels that, from an outside view probably lead nowhere, but just might lead to a golden trove. The confidence that allows me to plumb these tunnels is, from a certain perspective, irrational, but in expectation, it may make society overall more likely to have true beliefs.

In many ways, adopting this perspective has freed me from a poisonous dialectic. The man who wants to be right in everything he propounds is alternatively caught by the fear that he is wrong or by an overabundance of confidence he is right. Both unlimber you.

II. Why focus epistemic goals on the collective?

Why adopt this perspective rather than a more individualistic approach to rationality? There are really two questions here, viz:

What is intrinsically more important, society being right or you being right?

If society being right is more important than you being right, what's the best way, in expectation, to help society become more correct? How can we be sure it's not just each individual trying to be rational?

I think the first question almost answers itself. The second one is more interesting, I hope to treat it in more detail at a later date.

However, I would suggest that, as a preliminary sketch, every single individual trying to have the correct beliefs individually, and pursuing that in a rational way, would probably lead to an underinvestment in niche ideas and possibilities. Eccentricity ensures a diversity of ideas.

We also have to consider what I have elsewhere termed the paradox of the crowd. If you want to be right as an individual, your best bet is to adopt the most widely held beliefs,

on the basis of a wisdom of the crowds' argument- except in cases where there is reason to expect systemic bias. However, if everyone did this, the overall epistemic quality would go down. The crowds can only be wise if the individuals that make them up are not too over eager to follow the crowd's wisdom. (This is a bit like the problem of a stock market becoming overburdened with index fund investors).

III. Neglected truths

I previously suggested the two cardinal virtues for someone trying to help society get it right are being interesting and being honest. There's a third virtue as well, but it comes in at another stage of analysis.

Let's say that your aim is to help society find truths, but let's suppose that not all truths are equally in everyone's interests, so you start to think about what truths are, so to speak, likely to suffer from underinvestment.

I'm sure you'll agree with me that, at least one set of underinvested truths, are going to be those truths that, if uncovered, can help advance the position of the weak, at the expense of the strong (1).

So we come to another epistemic virtue that I call acting in service. Helping advance those ideas that are likely to be systematically squelched because they are inconvenient and difficult for those who hold power. You will note that while this virtue of service could be justified in non-epistemic terms, our argument for it here is purely framed in terms of the epistemic health of society: it is likely that ideas that disadvantage the powerful will be systematically neglected, therefore there is value in focusing our investigations there.

Utilitarianism is an egalitarianism

By utilitarianism, I mean the philosophy that holds that we should aim to maximize the overall welfare of people. Many philosophers working on political topics reject utilitarianism as a theory of the good society. This is fair enough, no theory will win universal approval. Most often, in the context of political philosophy, this criticism comes from the ostensible left of utilitarianism. Utilitarians are thought to be insufficiently egalitarian. The utilitarian, it is objected, only cares about the total and/or average amount of welfare. But surely the worst off have a special claim to our help, and surely inequality is objectionable in and of itself.

Positions that some egalitarian critics of utilitarianism hold include:

The difference principle: The difference principle is associated with John Rawls. It holds that between two societies, one can determine which is better by looking at the worst off person in each, and seeing which of them is doing better. This is not as counterintuitive as it may sound. Have not many people said, “judge a society by how it treats its worst-off”? Under this rule, inequality is justified only if it is ultimately to the benefit of the poorest.

Prioritarianism: Prioritarians hold that changes in a person’s wellbeing should count for more if their wellbeing is lower. An increase in a poor person's wellbeing counts for more good.

Egalitarianism: Egalitarianism is like prioritarianism but different in subtle ways. Egalitarians hold that less inequality in a society is inherently good, and more inequality is bad. Sometimes it might be acceptable to increase inequality, but the badness of inequality must always be considered.

Luck egalitarianism: Luck egalitarians hold that everyone must have the same opportunity to succeed. This may sound inegalitarian to you because many people say they believe in equality of opportunity without meaning it. A true believer in the equality of opportunity, the luck egalitarian will reject things like inherited wealth. Because we do not control many of our talents, differences in talent will also not be a legitimate source of differences in outcome. One of the few things that will be a source of legitimate differences in outcome is choices about whether to work or take leisure.

All of these philosophies have sometimes used utilitarianism as shooting practice. They have criticized it as inegalitarian and seen it as indifferent to claims of fairness.

I'm concerned by this strain of critique because utilitarianism is a very egalitarian philosophy. Utilitarianism shares with its egalitarian critics an opposition to the status quo because the status quo is too unequal. What follows is a meandering collection of observations on equality, practical politics, philosophy, and utilitarianism.

1. The ideas that have real currency

It is odd that the debate between utilitarianism and more egalitarian philosophies has taken up so much space. Views on distributive justice popular outside academic philosophy are so much less egalitarian than utilitarianism.

In public discourse, a charitable reconstruction of the political center would be a qualified Nozickeanism. Nozickeanism here means that individuals have a right to their property through the history of the way they acquired it. Because it is a qualified Nozickeanism, proponents concede that these rights to property may be abridged under circumstances of great moral urgency, but only in such circumstances.

But stripping back the gloss of charitable reconstruction, I suspect the philosophy of distributive justice held by many is a philosophy that we will call rulesianism.

Rulesianism is a kind of folk philosophy held by many in the media class. To the best of

my knowledge, no professional political philosopher has ever been a rulesian. However, rulesianism holds a magnetic power over some segments of the public. The rulesian holds that if you gained your wealth through the rules, then by those rules, you own that wealth. It would be cheating to change those rules now. The wealthy winners played the game and won- that's the end of it.

Desert theory supplements rulesianism. People deserve their property because they have displayed good character to get it- thrift and industriousness. Yet we can see that rulesian considerations are more fundamental than desert considerations. This is because Rulesians say that even the nakedly undeserving get to keep their property if they acquired it following the rules.

One might think that rulesianism is just respect for incentives. If we don't let people keep what they gained through the rules, why would they bother to gain anything? Society would be poorer as a result. Yet rulesianism goes deeper than respect for incentives. Granted, in a market economy, some inequality is necessary to create incentives. However, the rulesian seems more concerned with letting people keep what they hold than appraising overall incentive structures.

The obvious objection to Rulesianism is that the same rules that assign wealth also allow for wealth redistribution for the common good. Perhaps there is a way to make Rulesianism survive this objection but I can't think of it. I think philosophers must grapple with rulesianism despite its weaknesses- because it is a barrier to the egalitarianism many philosophers favor, and because political philosophers have an obligation to engage with popular ideas.

Moving along from rulesianism. If we go outside of public discussion and to the circles of technocrats, many fear that we are ruled by "government house utilitarianism". The reasoning seems to be that since economics is the queen of the social sciences in the opinion of government decisionmakers, and economics is utilitarian, utilitarianism

must be the ruling ideology of technocracy. This seems to be the fear of the Mouvement Anti-Utilitariste dans les Sciences Sociales- a group of scholars convened to fight what they see as the threat of technocratic utilitarianism.

Rumors of the utilitarianism of the economics profession are exaggerated. Since Pareto and Lionel Robbins, economists have usually rejected interpersonal comparisons of utility. To interpersonally compare utility is to make a judgment between two people about who has the greater welfare. Without the interpersonal comparison of utility, utilitarianism is impossible (I give my own approach to solving the comparison problem here). Instead of utilitarianism, an ideology we will call efficientarianism reigns among economists. According to efficientarianism what matters is to make Kaldor Hicks improvements. Intuitively speaking, to make a Kaldor-Hicks improvement is “to grow the economic pie”. This is to be done with no regard to how that pie gets distributed. The theory is that someone else will take care of distribution through tax and transfer.

It's not so much that economists believe in efficientarianism- many are quite sincere in their desire to see redistribution through tax and transfer down the road. Rather, efficientarianism is the practical result of what economists do. One of the problems with the idea that technocrats can get on with growing the pie and someone else will take care of redistribution in the future is that rulesianism means that people think that once you have an asset, you get to keep it. Efficientarianism among public officials and rulesianism among the political chattering classes combine to create inequality.

So rulesianism is in vogue among the media class and efficientarianism is often the result of economists' practice. It's odd then to see philosophers criticizing utilitarianism for being inegalitarian when it is so much more egalitarian than these ideas that have real power.

2. CBA and the misleading way we talk about “efficiency versus equity”

Part of the confusion around utilitarianism comes from the way we often talk about efficiency versus equity. This confused way of talking allows efficientarianism to pass itself off as utilitarianism.

Cost-benefit analysis is a form of economic analysis which attempts to quantify all costs and benefits in monetary terms. If you look at many documents which outline cost-benefit analysis, they will include a disclaimer like this: "cost-benefit analysis is only concerned with efficiency but not equity".

What they mean by this is that cost-benefit analysis (CBA) treats a dollar as a dollar in a peculiar way. Let's suppose the government is deciding where to build a bridge. There are two options. The first option, favored by Monoply Man Von Moneybags (MMVN), is to build the bridge at site A. The second option, favored by a village full of people, is to build it at site B. MMVN wants it built at site A because it will allow him to cross the river more quickly to get to his 117 hole private golf course using only a limo, not a chopper. The village all wants it built at site B because they all have ailing mothers the bridge will let them visit.

All 50 people who live in the village would be willing to pay 1000 dollars each to have the bridge built at location B, for a total willingness to pay of 50,000. 1000 dollars may not be very much, but it represents pretty much all the income each of them can spare. MMVN would be willing to pay 100,000 dollars to have it built at site A. The method of unweighted CBA suggests the government should build the bridge at site A.

You probably find this result disturbing (I crafted the example that way). But I want to focus in on something interesting about how we talk about this kind of CBA.

Economists call the result efficient and say your concerns with the result are equity

concerns. Economists grant equity concerns are valid. However, they argue that it is not the domain of economists to work out equity concerns, but instead political decision-makers. In practice, I suspect the politician, left with the clear CBA on the one hand, and unquantified equity concerns on the other, will do what the CBA says.

Here's my concern. I disagree with the recommendation of that cost-benefit analysis in the example I made up, but I don't think this is because of equity concerns. When I think of pure equity concerns, I think of the philosophies I described in the introduction that aims to be more egalitarian than utilitarianism

But you don't have to be a prioritarian or an egalitarian to disagree with cost-benefit analysis. If you are a utilitarian you can have another kind of efficiency concern. Distributive efficiency. Wikipedia defines distributive efficiency as follows:

“In welfare economics, distributive efficiency occurs when goods and services are received by those who have the greatest need for them. Abba Lerner first proposed the idea of distributive efficiency in his 1944 book *The Economics of Control*.”

i.e. in a situation of perfect distributive efficiency, each good and service is assigned to the person who will most benefit from it, which is to say receive the greatest boost in utility or wellbeing from it.

The danger in treating a dollar as a dollar in cost-benefit analysis is not just that it's unfair. The danger is that from a utilitarian perspective it's an inefficient way of assigning dollars or in-kind benefits to maximize welfare.

But an equivocation of our public dialogue- between efficiency in the sense of the economist, and efficiency in the sense of the utilitarian, leads us into confusion about utilitarianism. Since pure equity concerns do not directly play a role, many assume that

utilitarianism only cares about efficiency in something like the economist's sense. In truth, utilitarianism quite egalitarian, due to a concern for distributive efficiency.

3. Perhaps maximalist demands for egalitarianism can be less effective than strong but quantified demands for egalitarianism

Consider two separate claims one might make.

Our number one priority is to help the poorest people in society. This takes priority over helping any others, no matter the relative magnitudes

The income elasticity of the marginal utility of income is 2

The first, a gloss on the Rawlsian difference principle, allows no balancing whatsoever between interests.

The second is a little harder to unpack. What it means is that for every one percent income increases, the marginal utility of income falls by 2%. What would this mean in practice? Well, it would mean that an extra dollar going to a person on 25,000 dollars a year is worth 16x more than an extra dollar going to a person earning 100,000 year. It would also mean that if we have the choice between giving 7 dollars to a person making 25,000 dollars a year, or 100 dollars to a person making 100,000 dollars a year, the former is preferable.

Technically, claim 1 is infinitely more egalitarian than claim 2. However, I think there's a practical sense in which claim 2 actually turns out to be more egalitarian than claim 1. How?

When you tell people that X is infinitely more valuable than Y, they tend to just treat X as much more valuable than Y. This is why it's so dangerous just to say "lives are

infinitely valuable". In practice, no one will ever act as if lives were infinitely valuable- if we did, at the very least road usage would be much more restricted. So when you say life is infinitely valuable, you can actually diminish the concern people have for lives in practice.

If, however, you tell people lives are worth a very large but finite amount, say 10 million dollars, then it can enter decision-making functions. A strenuous effort will be made to avoid it. It is my sense that something similar might be true about the value we place on the welfare of the poorest. If we say "the position of the worst off is all that really matters" people aren't going to take us literally. Thus claim 2 may be, in practice, more egalitarian than claim 1.

Am I right about this? I don't know, but it is one practical reason to consider utilitarianism over the difference principle in politics.

4. Utilitarianism is so supportive of wealth redistribution that there is little practical difference between utilitarianism and putatively more egalitarian philosophies

There is a cottage industry of researchers that study the implicit eta or rate at which marginal social utility declines in income that is implied by the tax and transfer systems of different countries. That is the eta value that would be necessary to make those tax systems rational if the aim of the tax system is to maximize social welfare. Bargain et al. (2014) arrive at an estimate of inequality aversion for a number of European countries. This includes the Danish tax and transfer system which they find has an implicit eta of 3. They suggest that such a value is clearly beyond utilitarianism and that Denmark is effectively following the min-max rule of Rawlsianism: improve the position of the worst-off at all costs. An eta of three implies that an extra dollar for someone earning 30,000 is worth 37 times as much as an extra dollar for someone earning 100,000 dollars a year. You might think this could not possibly be a utilitarian approach.

Oh contraire. Utilitarianism can endorse what appears to be an η of 3 in the tax and transfer system- or even higher. My colleague Latty has shown this through work on wellbeing, diminishing marginal utility in income, and relative income effects.

A relative income effect is an effect on your welfare caused by your neighbor's income. Generally speaking, people desire to have greater, or at least equal income, than their neighbors. In very unequal societies, relative income effects tend to reduce overall wellbeing. Relative income effects are often taken to represent envy, but actually, they may represent all sorts of things. To pick an example at random, suppose you were poorer than many of your neighbors. This might make it hard for you to partake of the social life of your neighborhood. You may also find that only luxury goods and services are sold in your area, pricing you out of many things.

Evidence suggests that the effects of relative income on happiness are large. Relative income effects are another factor, in addition to the declining marginal utility of income, that pushes the utilitarian to prefer a more equal distribution of income.

Technically relative income effects should be treated as a separate parameter to η in models of the inequality aversion of a tax system. However, if relative income effects are not accounted for separately, as in Bargain 2014's model which does not include relative income effects, the apparent η implied by a tax system that is trying to be at least approximately utilitarian could be north of 3.

My colleague Kieran Latty has done mathematical work on prioritarianism and utilitarianism and found that using empirically plausible values of the declining marginal effect of income on wellbeing, as well as plausible estimates of the magnitude of relative income effects prioritarianism and utilitarianism gives almost the same result in almost all cases. Utilitarianism is so powerfully affected by these considerations that it is, in practice, not very different to prioritarianism (even when the degree of priority is

raised very high) and other forms of egalitarianism. Latty finds that many popular forms of welfarism converge in practice, drawn towards egalitarianism by the declining utility of income and relative income effects.

One philosophically interesting aspect of this is that it is, in part, the very same human intuitions of fairness that drive philosophers to egalitarian philosophies that make utilitarianism more egalitarian. People tend to flourish more without the negative effects of income disparity at least partly because they find those conditions fairer. Thus utilitarianism delivers the same result as egalitarianism for the same reason that many philosophers end up believing in egalitarianism- because a sense that equality (or approximate equality) is fair is built into us. Fairness intuitions are in a very direct sense, incorporated into the utilitarian calculus, not discarded by it. Another way to put it is this: fairness is built not into the philosophy itself, but into the people that utilitarianism is applied to. In turn, utilitarianism is responsive to the needs and wants of people.

5. How can philosophy fight when the enemy isn't ideas, but entrenched power structures? I don't know, but being clearer on what we're united around couldn't hurt

My argument, then, is that the fight between egalitarianism and utilitarianism is not as pressing as some have made it out to be. One objection that could be mounted to the argument I've made here is as follows. "Bear, you've said that utilitarianism is more redistributive than the actually existing political center, and egalitarians and utilitarians should focus on their common ends. However, we are doing political philosophy, not politics. In political philosophy, we argue out differences on the basis of their theoretical, not their practical significance. There is still a debate to be had between utilitarianism and more egalitarian philosophies, even if, in practical terms, they have more in common than is often recognized."

I acknowledge this objection. I agree there is a debate to be had. But I also think there is another project that is worth engaging in- the philosophical articulation of an overlapping consensus. An account of how different starting points arrived at the same conclusion. Meditation on how different starting points inform and reinforce the egalitarian case overall, and each tributary argument that flows into it. We might call this project of seeking and creating overlapping consensus on policy questions applied political philosophy due to its similarity with the project of applied ethics.

6. It's often forgotten that utilitarianism doesn't come with a specific theory of the good life

There's a sense in which utilitarianism, as an approach in political philosophy, can be agnostic on what the good life actually is.

According to utilitarianism, goodness is equal to the sum or average of welfare. Exactly what welfare is isn't specified by utilitarianism itself, but only by its variants (hedonic utilitarianism, preference utilitarianism, etc). One possibility, often neglected, is that utilitarianism can come with a rich theory of what the good life is. For example, utilitarianism can encompass a eudaimonic or objective list account of the good life.

The interesting feature of different concepts of welfare is that such concepts, as Hausman has pointed out, invariably intercorrelate closely. Measures of flourishing or capabilities correlate closely with measures of pleasure. Measures of pleasure correlate very closely with measures of life satisfaction.

This feature of human wellbeing means that, from the point of view of practical politics, it often matters very little what theory of the good we plug into our utilitarianism. Thus it is quite possible to be a political utilitarian, but not endorse any particular theory of what the good life is. That is to say, these concepts of the good mostly just track each

other anyway, so you can be agnostic about which one is the true measure of what matters since from a policy perspective, they amount to much the same thing- get one and you'll get the others.

One of the things particularly philosophers, who pride themselves on rich inner lives etc, find frightening about utilitarianism is the possibility that it will elevate the pig to the level of Socrates. The good news is that if you prefer a richer theory of the good life, which includes things like the full development of natural capacities, etc., etc, from the point of view of the political advocate it is unlikely to matter. The same priorities like poverty reduction, incarceration reform, etc. remain the same, whatever theory of wellbeing you plug into utilitarianism, and the same policies will almost certainly lead to all three of desire satisfaction, pleasure, and the full flourishing of human capabilities.

For the record, personally eudaimonic utilitarianism is my personal preference because it seems to me to do justice to the idea that life should be rich and full (it rejects wireheading), while still capturing what is persuasive about consequentialism, and the welfarist notion that what matters ultimately is the welfare of people.

Interlude: Colander

Colander argues that all of early utilitarianism was vague about exactly what the good was in a slightly different, but still interesting way:

What was meant by “good” for the question at hand was not expected to be fully defined or known beforehand. It was to emerge from the normative discussion. For some policy debates, material welfare might be the relevant goal and what was meant by good could be defined relative to standard economic measures. For other debates, “good” could mean freedom of choice and have little concern about material welfare. “Greatest good for the greatest number” was a malleable concept that would be defined by the context of the policy question, not by any fixed definition. Strictly scientific economic models

could not capture the many dimensions of “good,” which was why their scientific models did not lead to any policy implications on their own.

7. **Can utilitarianism be a critical philosophy? Yes.**

We often imagine the utilitarian as a decisionmaker, perhaps a bureaucrat, sitting alone in a tower overlooking those proverbial train tracks, making the decision whether or not to divert the train. But despite phrases like “government house utilitarianism”, the utilitarian is unlikely to be inside the tower, the utilitarian is much more likely to be sprinting up to the tracks trying to save the five, after the government made the call not to kill the one, because he was a citizen or his family were swing voters, or above all, because he had greater willingness to pay.

We too easily forget that utilitarianism as imagined by Bentham was a critical philosophy, intended to make the defenders of existing legislation, moral codes, and institutions nervous. Utilitarianism is a criterion for finding special interests, entrenched elites, and unnecessary cruelties. I suppose I see my (partial, qualified) utilitarianism in this spirit. It is not a way not of trying to create a closed ethical system. It is certainly not the last word on ethics. But it is a starting point for critical ethical discussions about the status quo.

Philosophers have a justified skepticism of unprincipled theoretical eclecticism which says “take a dash of this and that” but doesn’t explain how this and that can go together. Am I making an apology for such an unprincipled eclecticism here? No, I don’t think so. All one must grant in order to use utilitarianism as I’ve described is the following:

If, of a set of mutually exclusive and exhaustive alternatives, X increases welfare the most, this gives us strong reason to favor X.

Call someone who accepts the above a weak utilitarian. We are almost all weak utilitarians. Weak utilitarianism is all we need to embrace utilitarianism as a critical

yardstick for evaluating society. It is useful as a yardstick because it is very easy to quantify compared to many alternative approaches to the common good.

8. Utilitarianism as a public philosophy

The philosopher Goodin has an interesting discussion about utilitarianism as a public philosophy. By this, he means utilitarianism is as an ethical philosophy especially suited for public, rather than private decision-making. I agree, although I perhaps see it in slightly different terms. We can see utilitarianism as useful as an approximation of the correct approach to politics without being utilitarians ourselves in our day-to-day ethical existence.

Why do this? Because many of the most plausible counterexamples to utilitarianism involve methods and situations which are not relevant to governance or, at any rate, are unlikely to be relevant to governance in a modern democracy. Won't utilitarianism lead us to wirehead the whole world? Doesn't utilitarianism lead us to neglect what is special about our friendships and individual commitments? All reasonable objections, but objections unlikely to arise in the context of working out the ideal progressivity of the tax system, or penal reform.

The point converges with that in the previous section on utilitarianism as a critical philosophy. We should not let counterexamples that are not immediately relevant lead us to discard utilitarianism as a measuring stick. There are merits to using utilitarianism in policy analysis- because it is so easily quantifiable-, even if one is not a utilitarian?

This leads me to think that we should see political utilitarianism as a distinct philosophy from utilitarianism generally. The political utilitarian also need not hold any particular theory of the good life (see section 6), and can be agnostic or even negative on utilitarianism as a system of ethics.

An agnostic philosopher considers God, the limits of reason, and universalism

I'm an agnostic. I wish I believed in Christianity (or at least a certain kind of Christianity) but I don't. I simply don't believe in God because I do not feel I have the evidence he exists. It's a marvelous idea, but where's the proof? But I've been thinking about God a lot lately as I am studying medieval philosophy, and it seems to me that many Christians have not fully grasped the implications of an all-powerful, all-loving, and all-knowing God, just how wonderful, magnificent, and beyond comprehension a thing like this would be. In particular, I've been bought back to thinking about God through a kind of aesthetic, philosophical wonder at the doctrine of universalism, that all people will be saved. Though I can see no reason to believe in the Christian God, universalism makes me sad about that. I thought I'd present some meditations on God that I've been contemplating, in the hope they're of use to believers, unbelievers or both.

{Perhaps this essay is a symptom of the fact that people are always telling me I write like Søren Kierkegaard and I'm leaning into it. Oh, dear.}

I- Prayers for the dead

Let me illustrate with an example. There is an argument between (most) protestants and Catholics about prayers for the dead. Catholics argue from a number of premises which include tradition, and a passage 2 Timothy 1:18 in which Paul appears to pray for a man named Onesiphorus who, based on contextual clues, maybe dead. Protestants reply by rejecting church tradition as authoritative and claiming that the contextual clues around the passage in 2nd Timothy can be read in other ways. Protestants also give as a master argument something like the following:

1. The apostle Paul would not do that which is futile.

2. If Onesiphorus is dead, he is either in heaven or hell.
3. If he is in hell, prayer is futile for he shall remain forever in hell.
4. If he is in heaven, prayer is futile, for he is already in a state of perfect and permanent bliss.
5. Therefore, if Onesiphorus is dead, praying for him is futile.
6. Therefore, either the apostle Paul was not praying for him, or he was not dead, or both.

Catholics simply deny premise 2, for they hold that many persons when they die go neither to heaven nor to hell, but to a realm called purgatory, where through punishment they are made perfect until they can go to heaven. In this state, prayers are meant to be efficacious for speeding a person's acceptance into heaven. Catholics have a number of reasonable biblical arguments for their position, and the protestants have a number of arguments for their own view, it gets complex.

But even if you don't deny premise 2, it's possible to hold that prayers for the dead have value.

In the Christian tradition, it is commonly held that God is outside time, or that, at the very least, he has a perfect knowledge of the future. Now suppose that Onesiphorus was dead, but that Paul didn't know his ultimate fate- he didn't know whether Onesiphorus was in heaven or hell. If Paul prayed for Onesiphorus after he died then God would know and have known for all eternity that Paul was going to pray for Onesiphorus after he died. In whatever mysterious way prayer entreats God to do good, he would thus take it into account before Onesiphorus died.

And why can't an omnipotent and omniscient God grant a prayer before it is made? Thus it seems to me that even if the protestants are right about purgatory, they are wrong to think that prayers for the dead have no merit.

Now I'm not claiming any special genius for myself, anyone who spends way too much time thinking about time would be able to come up with that argument, but I can very easily imagine an unsophisticated fellow who, upon learning that his brother died, begun to pray for him, reasoning that God had told Christians to pray for what they wanted. The local theologian scoffs, giving the argument I outlined above for the futility of prayers for the dead. The simple man replies that the salvation of his brother would be a good thing, and God says to pray for good things, so he should pray for it.

I think the unlearned man has the right of it here, and as long as the ways of God are extremely complex and mysterious, which the Abrahamic conception of God requires that they will be, he's always going to have the right of it. Much as I wish it were, it's not my religion, so I'm hesitant to comment, but I suspect a lot of Christians are missing out on the best part of actually being Christian- a sense of total trust even when you can't reason out a way things could work out.

II- Universalism

I bring up this seemingly abstruse argument to make a point. This is a possible way that God could bring about good that many people have probably never even considered because it's so far outside the ordinary human experience of time.

I was once told a story by a dear friend, a Jew, who attributed this story to a Muslim friend of his. I apologize if I repeat the story with any errors.

Once a sinner resolved that he would repent to Allah. He wanted to do it in the right way, so he decided he would do it upon reaching a certain city. On his route, the man dropped dead.

Allah put the question to his angels what the fate of the man should be. The angels discussed among themselves and came to a consensus that if he had made it halfway in his journey, he should be admitted into paradise, but if he had made it less than half the way, he should go to Jahannam (hell).

Allah agreed. As it turned out the man had been less than halfway to the city in which he planned to repent. Rather than sending the man to hell, Allah in his wisdom rewrote both history and geography to move the location of the city closer to where the man had died. The man was then admitted to paradise. The point of this lovely story as I take it is that attempts to circumscribe divine mercy are futile.

To be a Christian (or any kind of monotheist) is to believe in a being who can do good in an infinity of ways that you do not comprehend, and even an infinity of ways you cannot comprehend and who will do exactly that.

The only possible response to this is a state of perpetual, child-like hope and complete trust. “God’s ways are not your ways” or “What is impossible for man is made possible with God”. It’s all going to turn out not just well but in some sense, as well as it possibly could.

Let’s take another example. There is a long-standing debate between those who think that the bible teaches that every human will ultimately be reunited with God in heaven and the more common view that a lot of people, perhaps even a majority, perhaps even a vast majority, will spend eternity being tortured in hell. The philosopher Keith DeRose gives an excellent summary of the universalist case here. My conclusion, on studying the texts pretty closely, is that if I were a Christian, I would throw my hands up in despair of finding an answer, but hope that everything would turn out well.

On purely textual grounds, I would be pretty comfortable arguing it either way. Pretty much the only positive conclusion that I would be willing to draw from the seemingly contradictory texts in the bible on this subject is that something strange, perhaps beyond the human power to comprehend, but certainly beautiful and wondrous was going on. I think the theologian Hans Urs von Balthasar contention that Christians have a basis for hope that universalism might be true, but no basis for certainty is a sensible position.

But again, I got to thinking about all the spanners that thinking through the implications of belief in an omnipotent, omnibenevolent God would throw in the works of trying to figure out the answer to this question.

One argument against Christian universalism works as follows, viz:

1. There are many passages that establish a belief that Jesus is God, and died for the sins of all is necessary for salvation.
2. There are many passages that establish that once you die, that's it, there are no second chances.
3. Some people die not believing in Jesus.
4. Therefore some people die without salvation.

It's not clear to me why, if you believe in a genuinely omnipotent, omniscient God who loves us dearly, but has zero respect for our ordinary conception of how things work, you should accept premise 3. I think there are at least two alternate possibilities:

How do you know anyone has ever died not accepting God?

How do you know any unbeliever has ever died on this earth?

It would be no problem, whatsoever, for God to stop time at the moment of an unbeliever's death, and right before they die have a conversation with them that convinced them to convert. But we needn't imagine time stopping either. It would be just as easy for God to whisk away unbelievers at the moment before they die, replace their bodies with an artificial corpse, heal the dying unbeliever, instruct them in faith, and then let them pass on.

(Two interesting facts about this bizarre idea that maybe people aren't actually dying, viz 1. To the best of my knowledge, the only individuals definitively stated to have died in the New Testament after the death of Christ are Christian 2. It may help make sense of another strange and seemingly false passage: "Truly I tell you, some who are standing here will not taste death before they see the Son of Man coming in his kingdom.")

The main objection I can see to the above arguments is that it's improbable- why believe in this sort of thing without evidence?

My response is that it's improbable only if you haven't truly internalized the idea of an omnipotent, all-loving God. We think of things like this as improbable because it would be impossible for us to do anything like this. If God exists, it's as simple for God as it is for you to raise your hand, in fact, it's infinitely simpler.

That's what I'm trying to get across with this essay. If you really accept and submerge yourself in the idea of an all-powerful, all-loving God then the way things are for you is really like a two-year-old child who has only just learnt to speak sentences trying to understand their loving parent's role in the world. These fantastic possibilities are presented not so much as a way of understanding, but as a way of showing that even at the limits of our understanding, the possibilities get bizarre, and of course, we can't even imagine what is beyond those limits.

St Julian of Norwich was a medieval mystic. At one point in her life, she became deeply distressed because so many people were going to go to hell. As I understand it, she had a vision in which God basically said to her “I won’t explain all of the details to you, but trust me, it’s going to be okay.” At one point Jesus says to her:

“All shall be well, and all shall be well, and all manner of thing shall be well”

Julian of Norwich presented herself as unlettered and unlearned, confused by her own revelations, but simply trusting God that all would be alright, in the same way, a young child who cannot understand why his parents have to go to work but takes their word for it, trusts her parents. I want to suggest that not only is that perfectly reasonable it is the only reasonable response to fully internalizing the idea that God can do anything whatsoever, is not limited by our understanding or imagination, and wants what is best for everyone. Far from being a reflection on her unlearnedness, Julian of Norwich was one of the few people who actually got the rational response if you accept the premises of an all-knowing, all-loving all-powerful God. The suspension of judgment and total trust.

An introduction to Jesus of Nazareth considered as an ancient moral philosopher

In this essay, I summarize the views of Jesus, considered as an ancient philosopher which an educated person should know something about.

I'll start by laying my cards on the table. I grew up Christian, but I left the faith relatively early.

I'm coming to this from a state of despair. The point that "haha, many people who claim to follow Jesus really do no such thing", has been made many times, but today I saw something that made me snap. There was a debate about the following situation. A man was helping raise a child, and the child viewed him as his father. Then, a paternity test showed that the man was not his father. Did the man have an obligation to continue to help raise the child? A guy who had an Eastern Orthodox icon of Christ as his profile picture was calling everyone who said yes a "cuck".

Here's the problem with that. A lot of what Jesus said is written down. It's pretty coherent. Jesus was a specific person, with particular philosophical views on morality, philosophy and theology. You can read what he said (or what the biblical Jesus said at any rate) very easily. He's not just an icon that means "whatever I agree with". Whatever you think about this sort of thing morally, it is, at the very least, poor scholarship to not check the primary sources.

In short, I'm sick of people using Jesus as a symbol for whatever they want to think anyway. Jesus was a sophisticated religious and philosophical thinker. His beliefs envisage a relatively simple, compelling, but horrifically demanding moral philosophy that should fascinate, inspire and frighten us.

In this essay, I want to consider the Jesus described in the bible as a moral philosopher. In an increasingly secular world, many of us know less and less of what Jesus said, and even those who have read what he said have tried studiously to misunderstand it. It's my hope to contribute to public education by summarizing his views in much the same way that I would summarize the views of, say, Socrates. He's an important ancient philosopher and so you should know a bit about him, in order to better understand the society he shaped.

Here's a bunch of things I will absolutely not be doing.

Considering Jesus as a Christian would, as the incarnation of the one God and savior of the world.

Considering Jesus as an atheist, seeking to disprove Jesus's claims. This essay aims to be entirely neutral on the divinity of Christ (something I suspect Jesus himself would have thought was impossible, e.g. Matthew 12:30- "he who is not with me is against me")

"Secularize" Jesus, discounting the theological aspects of what he said.

"Historicize" Jesus, seeking the real Jesus behind the stories.

Rather I want to take the biblical gospels, and consider them as works of philosophy in their own right. Essentially, I want to read the gospels as if Christianity were an obscure sect that had died out not long after the death of Christ, the gospels survived as our only record, and I were a historian of philosophy reading them as a philosophical (and religious) text.

It is my vain hope that this will stop some people from believing, or pretending to believe, very silly things about the biblical Jesus when several primary sources all about him are very easily available to read.

The moral philosophy of Jesus

I take the core moral teachings of Jesus to be as follows:

A) Moral obligations to God flow from love and as such should reflect a practice of love before specific rules.

B) Morality between people flows from very simple premises. 1. Value everyone's welfare as much as you value your own. 2. Everyone means everyone. This is framed in terms of loving each other. Love is, fundamentally, not defined as a feeling but defined as a willingness to sacrifice your own interests and even body on behalf of the other: "This is my commandment, that you love one another as I have loved you. Greater love has no one than this, that he lay down his life for his friends." John 15:13-14. (We should recognize that R Hillel, the esteemed ancient Rabbi said something similar, replying to one who asked him to summarize the law, he said that one must love the Lord God with all your heart, and love others as yourself and that all the rest of the law is commentary)

C) God has various ceremonial and religious requirements of us, but before all else, what God wants for us is to care for one another. Thus point A reinforces point B. "Woe to you, scribes and Pharisees, you hypocrites! You pay tithes of mint, dill, and cumin. But you have disregarded the weightier matters of the law: justice, mercy, and faithfulness. You should have practiced the latter, without neglecting the former. You blind guides! You strain out a gnat but swallow a camel." - Matthew 23:23-24

For the same reason Jesus is portrayed as healing others on the Sabbath, allowing his disciples to pick grain on the Sabbath etc.

D) Both these forms of moral obligation- to God and man- should reflect a logic of love, not duty. To truly love someone, one must do everything one can for the beloved. The chief signifier of love is, again, the willingness to make sacrifices on behalf of the beloved. If you are thinking in terms like "how much do I owe"- e.g., how much do I

need to do to be a good person you have already failed because someone who truly loves others doesn't make accounts like this, they are as quick to act for other's interests as their own.

E) Thus, morality is infinitely demanding, you can't "do enough" and just leave it, the value of your actions is equal to your degree of self-sacrifice on behalf of others and on behalf of God. You are obligated to sacrifice, at least potentially, everything. Give everything you have to the poor ("Sell your possessions and give to the poor. Provide yourselves with purses that will not wear out, an inexhaustible treasure in heaven, where no thief approaches and no moth destroys. For where your treasure is, there your heart will be also...." -Luke 12:33-34) and potentially sacrifice your own life (as above). This is simply what you would do, if you actually cared about other people as much as yourself, and the Lord your God with all your heart.

A corollary of the above is that since God loves everyone, he identifies himself with their interests. Thus any action you take against other people is really action against God, and any action you take for others is really action for God. "Whatever you do unto the least of my brothers, you do unto me". If you deny your brother water, you are denying God water, and so on.

F) Because we all fall so far short of the standard of morality- of loving each other and God totally-, all moral failures are arbitrarily bad, and we all make moral failures all the time. This is not to say that some sins aren't worse than others- Jesus complains about those who "strain out a gnat while swallowing a camel" after all, but trying to see oneself as a qualitatively less bad sinner than others is futile. See for example:

"To some who were confident of their own righteousness and looked down on everyone else, Jesus told this parable: "Two men went up to the temple to pray, one a Pharisee and the other a tax collector. The Pharisee stood by himself and prayed: 'God, I thank

you that I am not like other people—robbers, evildoers, adulterers—or even like this tax collector. I fast twice a week and give a tenth of all I get.’

“But the tax collector stood at a distance. He would not even look up to heaven, but beat his breast and said, ‘God, have mercy on me, a sinner.’

“I tell you that this man, rather than the other, went home justified before God. For all those who exalt themselves will be humbled, and those who humble themselves will be exalted.” Luke 18:9-14

G) This part you will probably know because Christians emphasize it a lot. Because our actions fall infinitely short, an infinite act of mercy is required to reconcile us with God. The interesting flipside of this, however, is that the very same standards that God holds which mean that we will fail, also mean that God chooses to extend us a chance. The necessity of our failure, and God’s mercy on our failures flow simultaneously from the reality that God’s moral code is infinitely demanding- that infinite demand falls both on us (requiring us to do the impossible) and God (requiring him to give us a way out).

H) You can’t really accept infinite grace while refusing to extend forgiveness to others. “For if you forgive other people when they sin against you, your heavenly Father will also forgive you” -Matthew 6:14

I suspect the reasoning here might be as follows. If you won’t forgive someone else, it’s because you regard their sins as fundamentally worse than your own. But that means you can’t have truly grasped the infinite magnitude of your own sins.

I) Our infinite failure means that seeking to make moral distinctions which elevate some as worthy and others as unworthy is a farce, and really serves as a club for some of the most immoral people in society to assert themselves as moral rulers. (E.g. in numerous passages Jesus attacks the Pharisees and scribes for using religious and moral ideas in

self-serving ways and try to escape the demands of real morality, and give themselves status over others.)

J) Because we are all obligated to do as much as we can, those who can do the most for others, but are not doing so, are the most culpable and the most morally reprehensible. Chiefly this includes the rich. (See the story of the Widow's Mite). By the same token, the poor are to be exalted. It is perhaps a little unclear on Jesus's philosophy why the poor should be exalted, as opposed to merely punished somewhat less than the rich, but that's what the text says. "Blessed are you who are poor".

K) True moral leadership has the inverse of the normal trappings of leadership, it means accepting indignities. "And whoever would be first must be your slave"- Matthew 20:2

Jesus as proto-utilitarian

Do unto others as you would have done unto you seems to me, fundamentally, to be an approximation of utilitarianism. Obviously, the slogan, as formulated, is vulnerable to counterexamples. What if you're a masochist? Just because you want to be beaten up does not mean that you should beat others up. However, it seems to me that "do unto others as you would have done unto you" is meant to approximate "do unto others as they want to be done unto themselves" in situations where you have no specific information about what they want. Like the Act Utilitarian- a utilitarian who holds that we should evaluate each action by its likely consequences on the welfare of others- Jesus thinks that our moral decision-making should be rooted in a fundamental concern for the wellbeing of others. Jesus attributes to God the same concern for the wellbeing of others, taking precedence over ceremonial concerns (e.g. healing on the Sabbath)

I actually feel a great deal of kinship in that, for Jesus, the heart of morality is concern for others, but he's a fair bit vaguer on formulating exactly what care for others means. Jesus proposes a heuristic (do to other people the same thing you'd want done to yourself) that works for a lot of cases, but falls apart in others. Similarly, I think all of

morality is about caring for the welfare of other people, but I'm not really sure what that welfare consists in.

Jesus had no time for honor culture or attempts to impress with aggressive masculinity

Jesus, unlike the odd character that made me write this post, had no time for interpersonal aggression, or assertion of one's real manliness:

“That whosoever is angry with his brother without a cause shall be in danger of the judgment: and whosoever shall say to his brother, Raca, shall be in danger of the council: but whosoever shall say, Thou fool, shall be in danger of hell fire.”

Note that his warning against anger is “without cause”, but there is no mention of cause or not in the case of one who calls his brother a fool. One could read this as suggesting that calling someone a fool endangers one with hellfire even if one has cause. Certainly I don't think Jesus was holding the mental reservation “unless they called you a fool first” as he said this.

The duty to forgive is infinite:

“Then Peter came up and said to him, "Lord, how often shall my brother sin against me, and I forgive him? As many as seven times?" Jesus said to him, "I do not say to you seven times, but seventy times seven." (we can take “70x7 here as meaning “an indefinitely large number)

Even if one is attacked in an insulting way- a slap- one has the obligation not to retaliate, and indeed turn the other cheek to be slapped. It is not even clear that

self-defense is permitted- for Peter was rebuked for raising the sword for trying to protect Jesus.

Nothing could possibly matter more than our moral choices

Linked inextricably to the idea that morality is infinitely demanding is the idea that our moral choices are also infinitely valuable. No prize, talent, capacity, object etc. could matter more than whether or not you use that thing for good, i.e. to meet the needs of your neighbors and for the glory of God.

“For what shall it profit a man, if he shall gain the whole world, but lose his soul?”

-Mark 8:36

“Provide yourselves with purses that will not wear out, an inexhaustible treasure in heaven, where no thief approaches and no moth destroys. For where your treasure is, there your heart will be also...”-Luke 12:33-34

By the same token, one who leads others away from good moral choices has placed himself in infinite spiritual danger: “If anyone causes one of these little ones—those who believe in me—to stumble, it would be better for them to have a large millstone hung around their neck and to be drowned in the depths of the sea.” - Matthew 18:6

Is Jesus a critic of exchange economies?

We know that Jesus had serious concerns about exchange economies, reflected in his constant condemnation of the rich and exhortations to give away property (numerous passages, absolutely unequivocal in their plain reading), rejection of commercial activity in the Temple (Matthew 21:12–13), and perhaps also (although I have mostly tried to stick to the gospels and not the later books), his early followers’ private property free lifestyle:

“The multitude of believers was one in heart and soul. No one claimed that any of his possessions was his own, but they shared everything they owned.” Acts 4:32

Where did this critical approach to exchange economies come from?

My sense is that Jesus thought the urgent obligation of every human being was to attend to the needs of every other human- a limitless obligation to serve the interests of others just as much as you serve your own interests. I think Jesus saw the organization of economic activity through money and property- favouring those with the means to pay- as inconsistent with the moral demands the needs of others make on us “Give to everyone who asks you, and if anyone takes what is yours, do not demand it back. Do to others as you would have them do to you”. Luke 6:30

To the extent that Jesus envisaged an alternative, I suspect it was making and giving things to meet the needs of others, rather than for exchange.

Jesus as political quietist

”Jesus answered, “My kingdom is not of this world; if it were, My servants would fight to prevent My arrest by the Jews. But now My kingdom is not of this realm.””

"Later, they sent some of the Pharisees and Herodians to catch Jesus in His words.

“Teacher,” they said, “we know that You are honest and seek favor from no one. Indeed, You are impartial and teach the way of God in accordance with the truth. Is it lawful to pay taxes to Caesar or not? Should we pay them or not?” But Jesus saw through their hypocrisy and said, “Why are you testing Me? Bring Me a denarius to inspect.” so they brought it, and He asked them, “Whose image is this? And whose inscription?

“Caesar’s,” they answered. Then Jesus told them, “Give to Caesar what is Caesar’s, and to God what is God’s.””

So you may have read the preceding essay, saw that I attributed ideas to Jesus like a rejection of economic exchange and a belief that moral ideas are used to buttress the power of self interested cliques, wealthy cliques and thought “You’re trying to tell us Jesus is a socialist- what a typical modern reader, trying to project such ideas backwards in time”. But no, I don’t think Jesus was a socialist.

Socialism is a method of organizing a political body, but Jesus is an apolitical thinker. Jesus thinks we should all, individually and right now, start acting like the needs of others matter just as much as our own, rather than implementing a political strategy to make that happen.

I don’t know why Jesus rejected politics like this. Maybe it was just to survive given the political situation of the time. However, I do not think so. I suspect he rejected politics because he saw morality as demanding we all behave perfectly (not that he saw this as practically attainable), and would see any idea of a political solution as an inappropriate attempt to substitute for that.

A final comment: On the homeless

Special comment because I have seen a lot of this lately. If you claim to be a follower of Jesus, but hate homeless people (I saw a trad “Christian” directly say that they hate homeless people and consider them worthless), the biblical Jesus implies that you will burn in hell.

“There was a rich man who was dressed in purple and fine linen and lived in luxury every day. At his gate was laid a beggar named Lazarus, covered with sores and longing to eat what fell from the rich man’s table. Even the dogs came and licked his sores.

"The time came when the beggar died and the angels carried him to Abraham's side. The rich man also died and was buried. In Hades, where he was in torment, he looked up and saw Abraham far away, with Lazarus by his side. So he called to him, 'Father

Abraham, have pity on me and send Lazarus to dip the tip of his finger in water and cool my tongue, because I am in agony in this fire.'

"But Abraham replied, 'Son, remember that in your lifetime you received your good things, while Lazarus received bad things, but now he is comforted here and you are in agony. And besides all this, between us and you a great chasm has been set in place, so that those who want to go from here to you cannot, nor can anyone cross over from there to us.'

"He answered, 'Then I beg you, father, send Lazarus to my family, for I have five brothers. Let him warn them, so that they will not also come to this place of torment.'

"Abraham replied, 'They have Moses and the Prophets; let them listen to them.'

"'No, father Abraham,' he said, 'but if someone from the dead goes to them, they will repent.'

"He said to him, 'If they do not listen to Moses and the Prophets, they will not be convinced even if someone rises from the dead.'""

Luke 16:19-31

Psychology

Harm OCD, a brief introduction

This article - and all my other articles dealing with OCD- contain discussions of very disturbing topics including mental illness, sexual assault, racism, pedophilia, sexism, extreme violence, etc. I speak about this stuff in brutal detail, because I think it's important we be clear on the nature of this illness and the kinds of distress it can create.

A reader asked me to explain more about my experience with OCD. I declined because I've already written a fair bit about it and feared I didn't have anything else interesting to say. However, some events on Twitter—the horrific spectacle of some faux-woke people saying to the mentally ill that they are just bad people, convinced me that an explainer about a form of mental illness called Harm OCD is necessary. At several points, I help myself to immoderate language. I want to be clear that this isn't because I'm being melodramatic, it's because I feel extremely immoderately about this topic.

1. How OCD and harm OCD work

OCD is a mental illness affecting 1-2% of the population. While it's not a competition, and I would never downgrade anyone else's suffering, there are signs that OCD can be among the most serious of anxiety disorders. For example, it typically must be treated with much higher doses of SSRIs than other forms of anxiety and depression. OCD is not a fun or cute disorder.

Wikipedia defines it quite adequately:

“Obsessive-compulsive disorder (OCD) is a mental disorder in which a person has certain thoughts repeatedly (called "obsessions") or feels the need to perform certain routines repeatedly (called "compulsions") to an extent that generates distress or

impairs general functioning. The person is unable to control either the thoughts or activities for more than a short period of time.”

Harm OCD is a form of OCD. In harm OCD the patient falsely fears that they will, or even that they have already in the past, do something destructive and unethical, opposed to their own values. Psychiatrists and psychologists have documented numerous instances of what we are calling harm OCD. Examples of harm OCD form an important part of diagnostic implements like the Y-BOCS. However, the impetus for thinking of it as its own unique category of OCD, with special complications and features, has largely been driven by the community of its sufferers.

To be clear, in current theory, harm OCD is not a separate illness from other forms of OCD, it's only a different kind of manifestation. It's not uncommon for people suffering from harm OCD to have experienced other types of obsessions and compulsions in the past, or even to experience them at the same time.

Some examples of harm OCD include:

- Susie is concerned that when she is cooking for her children she will include bleach in the dish and kill them.
- William is worried that he will sleepwalk in the night and stab his partner to death.
- Jiang is worried that she will grope and sexually assault a child while walking down the street. She fears she may have already done this and not remembered.
- Morgan fears that he will, absentmindedly “on autopilot”, write out a confession to a horrific crime he hasn't committed. He never throws out any paper in case he has accidentally written confessions on them.
- David is worried that he will just “flip out”, “lose control” and start beating up people as he walks down the street.

- Kiara is worried that she will mutter the N-word when she is on the phone with a black client under her breath.
- Lisa is concerned she will accidentally perform a hit and run. Every time she hits a bump in the road she reverses, to make sure it wasn't her running over someone and not realizing it.
- Trevon repeatedly has images pop into his head of him raping and murdering people. He worries this means he is a psychopath who wants to rape and murder people.
- Ayaan had sex with a woman several years ago. When he looks back on it he worries what if the woman wasn't really consenting? What if he pressured her into it, not meaning to? He replays tiny details of the event in his mind.

An attentive reader might be thinking "I see the obsessions, but where are the compulsions' ". Compulsions in this form of OCD can be subtle and less obvious than, say, cleaning compulsions. A sufferer might keep their hands in their pocket so that they don't accidentally grope someone, or mentally replay the events of the day, looking for evidence that they hurt someone inattentively. They could try to think good thoughts, to cancel out the bad thoughts. They may mentally rehearse arguments about whether some occasion on which they had sex was coercive or not. Forms of OCD where the compulsions are very well hidden are sometimes called Pure O OCD. Some people even believe it is possible to have OCD with no compulsions at all, even hidden ones—just obsessions—however this is controversial.

In addition to harm OCD there are two forms of OCD that are closely related and worth mentioning because they also integrally involve the values and identity of the subject. One is sexuality OCD, where the patient is afraid that they have a different sexuality than the one they usually regard themselves as having. Some common examples, A) a straight man worries that he is secretly a homosexual because he keeps having unwanted thoughts about having sex with men pop into his head. B) A man or woman

who keeps having unwanted images of having sex with children come into their head and worries this means they are a pedophile.

The other is scrupulosity OCD, where the sufferer worries about blaspheming against religion, or not being sincere in their faith, etc., etc. They may have, say, involuntary sexual thoughts about the virgin Mary and be deeply distressed by this. Martin Luther may have triggered the reformation, partly because he was driven to distress by Scrupulosity.

People with any form of harm OCD are not dangerous. Professor Ross Menzies once told me words to the following effect:

If hypothetically he wanted to be absolutely sure that he was not associating with someone who would perpetuate a hit and run, he would be well advised to spend time with an OCD sufferer afraid of doing just that. If he wanted to be absolutely sure his children would not be abused, he would leave them in the care of a person with pedophile OCD. If he wanted to associate with someone who was controlled and unlikely to lash out, he would associate with someone whose OCD made them terrified of exploding in anger, and so on.

He even suggested that harm OCD could be thought of as the opposite end of the spectrum to sociopathy, and disorders where there is an absence or mutedness of conscience.

The tragedy of harm OCD is that, through a process akin to natural selection, the mind picks the worst possible fear—the one most opposed to the sufferer's deepest values—to inflict on them. This is what sticks the best—what is hardest to shake and what will keep popping into the head. The sufferer is effectively being punished for the depth of their opposition to running people over, or raping children, or being racist, or stabbing their partner.

2. Twitter and greivous wrongs against the mentally ill.

Twice now I have seen rubbish on Twitter of the following form:

1. Someone says they have intrusive thoughts about performing a deviant act [the two examples I've seen: sexually assaulting a child, saying the N-word].
2. People pile on and say that it's disgusting that a person would admit to being tempted to such acts.

This is a horrific thing to do, and contemptible among those of them that should know better. It is one of the most hurtful things you could possibly do to a person with harm OCD, and their special vulnerabilities and fears mean that it really is no better than physical violence against them. If you only take one thing from this article, let it be this having an intrusive thought about doing x does not mean you are tempted to do x. Quite the opposite, it means that you find x especially abhorrent or terrifying. Many people with harm OCD may believe that they are tempted to do horrific things but they are not.

How can a person falsely believe that they are tempted to do something horrific? Well, there are many ways, but here is one example.

Try not to think of a pink elephant. You can't do it, right? Well, imagine if instead the thought you were trying to block out was serially raping and murdering women. You can't force yourself not to think about something. In the end, the thought pops into your head so many times that you begin to fear that it must be because you desire it. In truth, it's the exact opposite. Because you want nothing more in the world than not to imagine these horrific thoughts you can't stop thinking about them.

It should be obvious why telling a person with harm OCD that their condition proves they are dangerous or evil is one of the cruelest and awful things you can do to a person

with words.

3. Harm OCD as self-censoring illness

Harm OCD is a self-concealing illness.

Because it specifically involves things you are horrified of it makes you not want to talk about it. You fear that if others find out about these thoughts, you may be ostracized. Tragically, as the events on Twitter, I mentioned show, this can sometimes be true. I believe that the real incidence of people with harm OCD is probably much higher than we know, for this reason.

Finally, people who talk about their harmful OCD in detail, despite it representing everything they abhor, are heroes. They are risking their reputations and their mental stability to tell you something very painful about their experience. May God have mercy on your soul if you use that as an opportunity to turn on them.

Obsessive Compulsive Disorder and the origins of religion

Obsessive Compulsive Disorder (OCD) is well known to interact with pre-existing cultural and religious beliefs and practices. I was interested in coming at it from the other direction—is there evidence that culture, particularly religious culture, has been shaped by OCD?

Although I can't prove it absolutely, I believe there is enough evidence to make the idea intriguing, and worthy of further study by someone better suited than me. The idea that mental illness might be linked to the origins of religion is not new, but I have rarely seen the idea in the context of OCD specifically. However, I suspect that there is a connection, and the effect has been profound.

It is worth stating categorically that a link between religion and OCD is not intended to denigrate religion, nor OCD sufferers. As a lifelong sufferer of OCD myself, such an intention is the furthest thing from my mind. OCD is associated with many positive traits and traits which can be positive in some situations—caution, concern for our impact on others, cleanliness, seeking certainty, aiming to systematise etc. Although these traits tend to get out of control in OCD, at other times and parts of life they can be profoundly useful. If there is a relationship between OCD and the development of religion, the contribution of OCD could well be positive.

Nor does saying that OCD played a role in the formation of religion imply that great religious figures all had OCD (though many may have). Rather, ambient forms of OCD might have influenced religious thinkers and their interpreters. One might also believe that even people without clinical OCD can tap into an OCD like mindset, especially when concerns as vast as salvation and divinity are on the line.

Personally I see the link between OCD and religion as one of OCD and certain aspects of religion utilising the same 'circuitry', although outright cases of OCD most likely play a role in the development of religion as well.

The idea of a link between religion and OCD is far older than my poor thoughts (notably, Sigmund Freud championed it, although there is surprisingly little development in later literature), but I wanted to explore it in a more contemporary way. I've gone looking for information on people who have tried to develop a connection, but as best I can determine, most of the literature tends to be practical and epidemiological, with little of it coming from a studies-of-religion perspective.

What follows is my attempt to trace what I see as shared themes between OCD and religion. The exact role, if any, that OCD, and OCD-like mental states, played in the formation of specific religious beliefs is probably in most cases lost to time now, yet as we will see there may be grounds to suspect it was a rather large one.

Themes that OCD might have contributed to the development of religion:

In what follows we examine parallels between OCD and religious thinking, and speculate about the possibility that OCD-like modes of thinking may have contributed to their development in religion. These include: 1. Concerns about purity 2. Fear of offending the sacred 3. Rituals and ritualistic behaviour 4. complex, detailed and guilt driven forms moral reasoning, and 5. ideas of demonic or external possession. While many of the parallels we discuss are not entirely unique to obsessive compulsive disorder, the overall picture is of a striking resonance between the focused but fearful reasoning of OCD, and similar trends within religion.

1. OCD and similar mental states as an explanation for the origin of religious fears around contamination and purity

A common concept in OCD is the fear of contamination, although today this fear is often expressed in terms of the germ theory of disease, prior to the development of such it would likely have taken different forms.

Religious categories of ritual and dietary purity are often worked out with a logic that is at once rigorous and dreamlike, mimicking the ‘style’ of OCD thinking. This overlap in content (concern about purity) and style of thinking (internally coherent, deeply elaborate yet bizarre) is striking. There is a deep concern for how various unclean things might infect various clean things, of what is capable of ‘carrying’ or ‘transmitting’ uncleanness, about making sharp rulings in unusual and borderline cases and so on.

The book says do not mix a kid goat in its mother’s milk. Easy enough. But the book says don’t do it three times? There must be something extra here. Better not mix meat and milk, and keep a four hour gap between eating one and the other, just in case.

2. Fear of offending the sacred, blasphemy, scrupulosity and the origins of sacredness

A very common form of OCD is scrupulosity, a fear of offending the divine through uncontrollable sacrilegious thoughts, words, mental images etc, or through very small lapses in action or words that would seem unimportant to many. At first glance one would assume that the idea of scrupulosity presupposes the prior existence of a concept of the sacred, yet I can imagine ways in which symptoms of scrupulosity and a sense of the absolute sacred might co-develop. The belief that one must not even think—let alone say—certain things about a being, even in jest or illustration, elevates it to a higher level of sacredness. For sacredness to exist, the possibility of blasphemy must exist. The constant guilty struggle with one’s own thoughts and micro-behaviours about

the sacred can further drive greater religious engagement, and the development of cycles of repentance, perhaps contributing to the often dramatic rituals of self-abasing penance and the purging of sin common to so many religions.

3. OCD and the origins of religious ritual and ritual magic

Numerous OCD behaviours are rituals, in the sense of stereotyped series of actions intended to influence seemingly unrelated actions. Tapping a spoon three times to make sure your husband does not die on a ship, for example. Or Saying the Lord's Prayer 27 times to make sure that your children do not die of a horrible disease.

It is easy to see how, in a society which did not necessarily dismiss such rituals like ours tends to, someone might develop a belief about a necessary series of actions and this might be adopted by the society at large, especially if the ritual 'worked' several times. OCD even comes with a ready-made explanation of why rituals sometimes fail, since there is an overwhelming concern that one will slightly mess up one of the innumerable complex parts. Ritual failure is itself an important theme in many religions.

4. OCD, guilt and the development of categories of moral reasoning

A common form of OCD is a supreme concern that one will, or that one already has, acted in a way wholly dissonant with one's moral values. One form in particular is the fear of negligently or actively causing others harm, or violating the moral order of the universe, with actions that would, to most outside observers, look harmless.

Replaying past actions for evidence that one did the wrong thing in one's mind and endlessly vetting possible actions, could drive the creation of new categories used in moral and religious thinking.

To choose a few examples, categories of Islamic jurisprudence, Catholic moral philosophy and—perhaps above all—the Jewish Halakha—display an exemplary combination of deep caution, deep concern, deep subtlety and exceptional

systematisation. Relatively harmless or even completely harmless behaviours are condemned in the strictest language. St Thomas Aquinas claimed that masturbation was worse than murder, Jesus said that thinking about another woman sexually was a form of adultery and that calling one's brother a fool was worthy of the fires of hell, the Talmud says that it is better to be burned alive than to embarrass someone in public. The fusion of intense anxieties with careful legalistic thinking bears a weird resemblance to OCD, and the longer one reads the texts, the stronger the parallels seem. Neurological evidence suggests that OCD may be a disease of hypermoralism, and is deeply implicated in the over activity of neural pathways associated with moral reasoning and thinking (forming a possible inverse to antisocial personality disorder). It seems to me very plausible that inclinations towards hypermoralism may well be one of the motors driving the development of moral thought in a religious context.

But unreasonable rules give rise to resistance, and religion is no exception. Resistance to hypermoralism, by OCD sufferers who had become sick of it might form another motor driving moral development within religious traditions. An ongoing dialectic, sometimes creative and sometimes destructive, between the spirit and the letter of the law, is a theme in every religious tradition I have ever studied. I'm not just trying to have it both ways here—there's a least one very good case study of this kind of 'bending in the opposite way' reaction to hypermoralism by an OCD sufferer with a profound effect on the history of religion: Martin Luther, leader of the protestant reformation. Martin Luther is fairly well established to have suffered from scrupulosity, a religious form of OCD. Luther's continual sufferings, and deep fear that his behaviour was inadequate almost certainly contributed to the theological views which caused him to propose that belief in god alone was enough for salvation, thus attempting to cut the cycle of fearful hypermoralism.

5. OCD and demonic possession

While many mental illnesses can be seen as 'demonic possession' by those without a psychological background, the potential for OCD to be perceived in this way is often

forgotten. While I don't think the majority of cases of 'demonic possession' are really OCD in disguise, I suspect a substantial minority may be.

I would argue that many forms of OCD—such as uncontrollable thoughts of the violent, sexual or sacrilegious which go against the values of the sufferer could be seen in some societies as a sign of mental demonic harassment. Many sufferers of OCD (falsely) believe that they are experiencing barely controllable urges to do violence. Such feelings could very easily be mistaken for a sign of demonic activity, either by the sufferer, or by society at large.

6. Religious (Self)-Consciousness and internal struggle

But more importantly than extreme cases like demon possession there is the form of self-consciousness associated with many (though not all) forms of religious life, the sense of oneself as a tempted being—a being who is constantly in an unwilling dialogue with dark forces that wish one to sin. Whether these be conceived of as as internal, external (like a demonic tempter) or something ambiguous and intermediate (like the Yetzer Hara) the effect is the same, the self is seen as divided, and torn between obedience and sin

.

The parallel in OCD is found mostly in cases of morally charged OCD—scrupulosity, harm obsessions and sexuality obsessions. The sufferer can be confronted with a sense of division about themselves, about who one is, what one has done and what one might do—a kind of fragmentary consciousness in which parts of one's consciousness sometimes seem work with a vicious semblance of autonomy in order to perversely thwart the whole is fundamental. Many sufferers of OCD begin to conceptualise their OCD thoughts and impulses as a cruel and bizarre stranger living in their mind.

We might also point to the internal struggle between doubt and belief that plays such a role in the life of the believer, and such a role in OCD. Certainly there's a parallel in concepts here, although whether there's a deeper or historically important parallel is harder to say.

Summing up the historical case

You're a psychologist, someone comes to you and pitches a hypothetical. There's a patient who keeps strict rules regarding bathing, refuses to mix certain foods, repeats certain words at certain precise intervals throughout the day, regards certain behaviours as absolutely taboo for reasons which they either cannot articulate, or which seem bizarre, will only dress in certain specific ways, is afraid that they are a deeply wicked person and struggles with questions about their moral identity. If you had to take a guess at the diagnosis, what would it be?

Evidence of an association today

To bolster our historical associations, let us take a brief look at the very large literature suggesting that religion and OCD remain linked today. Degree of religiosity is linked to likelihood of developing OCD, this has been validated for a wide variety of monotheistic religions including:

Judaism,
Christianity
and Islam.

While formal research has focused on these religions (largely because of their popularity in the regions where research has been conducted) the internet is full of heartrending stories of people struggling with these problems in all kinds of religious contexts including Buddhism, Hinduism and Jainism.

There is little doubt that membership in at least some religious communities is associated with OCD. The direction of causation remains unclear.

Conclusions

What we've presented here is very partial and tentative. The idea of a strong connection is far from certain, but many individually modest pieces of evidence can form a greater image

.

I want to reiterate that it is not my intention to contribute to a crass reductionism here. We are not saying that major religious figures were all OCD sufferers. We're not making an argument that religion is a disease. Reasoning of the form "OCD might have contributed to religion, OCD is a pathology, therefore religion is pathological" is lazy and dangerous.

A more fruitful approach, and what I'd like to see, is greater investigation of the link between OCD, mental states which parallel OCD, and the development and survival of religions over time. Rather than denigrate religion, I would hope such a study might deepen our appreciation of the social role of both religion and obsessive compulsive disorder.

A final quotation:

"Luther's first years in the priory were thus a time of interior tension, spiritual struggle and suffering. The hopeless feeling that he was not numbered among the elect but among the reprobate overcame him and grew stronger as he grew more and more conscious that he did not fulfil God's commandments in all things. Since he began early to condemn as sin every movement of natural appetite, even though unwilling, and since, with his exuberant vitality, such movements kept recurring, he supposed himself to be full of sin, and no prayer, fasting or confession could free him of this terror."

Karl Adam, *Roots of the Reformation*.

Lessons I squeezed from a lifelong severe mental illness

In the past I've written both an essay on the philosophical aspects of what I learnt from OCD and another essay on the practical measures that helped me keep it under control. Now I wanted to convey the little life lessons I learnt from decades of chronic mental pain. Stuff that isn't directly about the disease, but might help anyone.

Plot friendship

Friends have kept me alive. I have rather a lot of them, despite not being that charismatic (or perhaps that's false modesty and I am charismatic, but if so, it was very hard won). You need friends. Chances are you will not get friends organically unless you are in school or maybe in college. So you must make it happen. That means taking deliberate and systematic action.

Do these things 1. Give yourself chances to meet people. 2. Keep a list of the people you want to know better. 3. Make a list of the friends you do have and keep an eye out that you're not neglecting any of them.

I know this sounds weird, maybe even creepy. However, we live in a newly atomised world—it was much less atomised even half a lifetime ago. The conventional wisdom hasn't kept up with the atomisation. A much more active approach to your social life is needed than TV and books would indicate. If that makes you feel a little bit like the talented Mr Ripley so be it. Plot friendship.

You can't enjoy what you grasp

You can't really live if you hold your life too dear. So many things I worried about losing through my fears—freedom, dignity, life, reputation—I could not enjoy because I spent so long fearing for them. If you hold onto things too tightly, you lose time with them just as surely as if you had thrown them away.

Music helps

One morning I felt so bad—trapped and cold. Then I turned on the *Mystery of Love* by Sufjan Stevens and I felt completely fine.

We all know this right? “Music can help with depression!” but it’s more than that. Music, poetry, visual art—we tend to treat these things as “nice”. In reality, these are load bearing pillars of our soul. You are cross-stitched with little bits of art and science and philosophy. That song lyric you can’t stop thinking about? It’s integrated into you now. Engineers and artists are the two closest things to sorcerers that exist. We don’t take them anywhere seriously enough. Cherish those who give you the materials to build yourself.

I regret every time I’m scoffed at someone singing along to music. I recently took up singing and can’t believe I didn’t do so years ago. Yousician costs 40 dollars a month and takes 10 minutes a day. Maybe my efforts are laughable, but it connects me with strips and veins of myself that would otherwise be inaccessible.

Life is better with a crazy ambition

I have a crazy dream. I won’t say “working towards that dream keeps me sane”, but it’s one of the little things that helped a bit. If you asked me what the odds of me pulling it off were I’d say about 1 in 30,000. That it is technically possible and that I want it to happen are about the only two things in its favour. I’m not going to reveal what it is—it’s

too embarrassing—just think of some juvenile quest and you’ll be in the right vicinity. I selected it on the basis that even if it fails it will help me achieve a bunch of valuable intermediate goals along the way. Consider picking one for yourself. Make sure it’s not dangerous to you or anyone else. If you’re not embarrassed by the thought of sharing it

publicly it's probably not crazy enough. Go wild, I give you permission to be a megalomaniac about this.

Whatever you do to the least sets the standards for everyone else.

My anxieties revolve around the fear that I might do something terrible. As such, I have spent far more time than most people contemplating what it would be like to be one of the most reviled people in society. The conclusion I keep coming back to is that the way we treat the most reviled people—think convicted serial rapists or killers for example—sets the minimum standard for everyone, and that has a trickle up effect on those nearish the bottom. It is in everyone's interest to set that standard higher, because none of us know what the future holds. It may be true that you will never be in quite such a reviled position, and it may be true that right now you are far from the bottom. However, anyone can fall low enough that the minimum standard will be one of your few safety nets. You could end up like one of the people in Ronson's book, for example.

Prayer helps even if you don't believe

I don't believe in God. I wish I did, but that's not something I can control. Nonetheless when I become depressed or anxious I often pray. Sometimes I pray for myself, but mostly I pray for some kind of universal redemption or outpouring of mercy. Sometimes I weep while I pray. Sometimes I feel angry, and ask him for answers like Job shouting into the whirlwind. Sometimes I protest like a Karen asking to speak with the manager of the universe, appalled at what the customer service has done to 100 billion humans. Then I reflect that if God- omnipotent, omnibenevolent- is real, then by definition he has his reasons. I have no idea why, but this process helps me.

I dreamt I saw

A portion of God's face

Only by his grace

Did I not see more
Lest I end like the sea
Ends a tributary creek

Rich in the mystery of darkness
Abundant in the revelation of light
The fire of his eyes is justice
And from that flame rises
The incense of mercy
And all the heat is love
And the cold is love too,
Lest we burn away
All there is, is love

Please.

OCD, mental illness and "cancel culture"

I want to use OCD as a lens to understand and critique what is sometimes called “cancel culture” and the discourse around it.

Fear of cancel culture is moving like an infection through the OCD sufferers. Why, and what does this tell us? Someone posted an interesting thread on the OCD subreddit recently:

“Does anyone else have an OCD-fuelled fear of being "cancelled"? Throwaway because this is a subject I'm mega anxious about!!! Basically, the title: Does anyone else have really distressing obsessive thoughts about being "cancelled" online/losing their livelihood etc. because of stupid things they did as a teen? I'm just petrified that people will find out about bad stuff from my past, they'll tell my employer, I'll lose my job and never find one again, my life will be ruined, that kind of thing.

When the thoughts hit they just cycle and cycle around my brain and I find it so hard to function because what's the point in doing anything if my life will eventually be ruined for things I regret doing? I have to go through all of these compulsive behaviors like Googling my name to see if anything bad comes up and seeking reassurance from loved ones just to feel slightly better, and even then the fear still remains a bit.

Am I totally alone in feeling this way, or have other people experienced it to? If so, how did you manage to overcome it?”

Some interesting comments on the thread:

“I've even stopped pursuing my dream of being a published author because I'm scared that it'll only lead to being "cancelled"...”

“i'm still pursuing that same dream but i'm gonna use a pen name.”

“Cancel Cultures themselves have probably made the same or similar mistakes as the person that they are canceling. Either they have forgotten about it or kept it to themselves.”

“I’m so happy I found somebody who understands this I’m crying”

“ALL OF THE TIME!!! Seriously! I always ask my friends if anything I did back then and even now is awful or worthy of being cancelled, etc, and they look at me like I’m crazy.”

“One thing that helped me was realizing that people have had these kinds of concerns across cultures and eras and found them very disturbing. I don't know why, but seeing 'reputation' on this list of the eight worldly concerns really helped contextualize my own fear and put it into perspective (I am just one person feeling afraid about their reputation).”

“Very much. I’m waiting for someone to unveil the person I truly am underneath and for me to lose everything. Like sometimes it’s from things I know I’ve done and other times it’s just a vibe that the end is coming for me.”

There were the dozens of comments along the lines of “wow, were you in my brain”, “yes, 100% this” etc.

Finally, I saw this comment, which was so good I’m going to draw special attention to it:

“I honestly believe that cancelling predates cancel culture... it's a part of a general belief system that people are disposable cogs in a machine that pervades our culture as a whole. Like in the past you could get 'cancelled' by major film studios for being gay, for example. It connects to the idea that if people are 'a problem' you can 'get rid of them'”

by firing them, locking them away, etc. to keep the core 'pure.' As OCD sufferers we're disproportionately impacted by all forms of purity culture, which has deep roots in our society connected to racism, homophobia, and so on.

Then this caught my attention:

“I'm a psychologist and this is becoming a super common theme for my clients”

So I did a bit of searching to see if it really does go beyond this. I found dozens of other threads about the fear of being cancelled on the OCD subreddit, many with dozens of commentators, picking one at random:

https://www.reddit.com/r/OCD/comments/hdup6b/cancelcallout_culture/

Then there were threads on many other forums

There were podcasts:

<https://www.fearcastpodcast.com/2020/07/14/real-event-ocd/>

/

And articles:

<https://carleton.ca/determinants/2019/cancelled-overcoming-the-fear-of-a-social-media-presence-in-a-growing-call-out-culture/>

<https://adaa.org/learn-from-us/from-the-experts/blog-posts/consumer/metoo-latest-ocd-trigger>

<https://cognitivebehavioralstrategies.com/ocd-in-the-age-of-metoo-revisited/#.YHdvSegzZPY>

This blog post by a sufferer is particularly good

<https://notmakinglemonade.com/myblog/2020/2/5/im-so-ocd-about-scrupulosity>

Etc.

So from research, It looks like a number of psychologists and psychiatrists are reporting a lot of OCD sufferers coming to their practices scared of being canceled, Metoo'd, or similar. Patients are also talking about it on the internet. I find this interesting and want to understand it, as a way of critiquing—and understanding—“cancel culture”.

OCD is an opportunistic pathogen as one blogger noted. There are a lot of OCD sufferers paranoid right now about the coronavirus. This phenomenon of people being scared of cancellation because it's in the news isn't surprising in some ways, but I want to dig down into what it can further tell us.

I find the category of “cancel-culture” a little frustrating because it focuses attention on celebrities. The etymology here is telling “cancellation”, coming from the idea of canceling a show, movie, book, or whatever due to a controversy. The old label of “callout culture” seems more adequate in that regard—less focused on big names and celebrities who can be, metaphorically and literally, pulled off the air. I care a lot less about some celebrity being canceled (although I do care) than I care about some poor nobody getting called out, screamed at, told to kill themselves, etc. on the internet. But the move from “callout” to “cancel” is, in some ways, more accurate. “Cancellation” captures the killer instinct inherent in the phenomena. All pretense to a moral corrective in the term “callout” is now gone.

The other interesting thing about the lingo of “cancellation” is that an ordinary person who fears being canceled is implicitly comparing themselves to a celebrity. I think they are both very right, and very wrong to do so.

They are right to compare themselves to a celebrity because the internet, particularly Twitter, Facebook and Instagram have turned fame into a matter of degree, rather than a qualitative distinction. I'm not famous, but if I desperately want 10,000 people to read something I write, I can make it happen. Given enough time and desire I could probably swing 100,000. The line between "a big account" and "internet famous" is extremely blurry, and so is the line between internet famous and real famous. This is not to say that there are important distinctions between famous and not here, but more than ever, they are quantitative rather than qualitative.

We are in the future that Andy Warhol imagined. We have our 15 minutes of fame, or at least the chance to strive for it. However, all that fame consists in is the bad bits of being a celebrity—the whiplashes in public opinion—with none of the cash or groupies, and with no agent to manage our PR.

Yet it is also very illogical to consider yourself like a celebrity in this sense. If you're an ordinary person, unless you get very unlucky (I'm not denying it's possible), no one cares if you said the N-word on a forum once when you were 15.

I'm not saying ordinary people are immune to cancellation. I once sat down to list everyone I personally knew personally IRL who had suffered a major reputational blow of some kind—from being accused of sexual assault to being exposed for some past act of racism. There were well over 30 people on that list. Some were pretty much as deserved as these things can ever be, others were tragic (an abuse victim accused of abuse by her abuser).

The truth, then, is complex. OCD, as it often does, has a cruel grain of rationality at its core. People are being cancelled. Somewhat randomly, sometimes with weak cause. The odds of it happening to you though, especially in a way that matters, are pretty small.

Who is vulnerable to shaming?

There's a quote I once saw on, of all places, a Magic: the Gathering card that has influenced my life ever since. "Those without a guilty conscience need one. Those with a guilty conscience don't". How does that apply here?

And just who is most affected by shaming online? I don't mean here just people who get explicitly "canceled"—the celebrities etc. I mean people who are, in one way or another, humiliated, shamed, defamed, or otherwise caught up in the rough and tumble of online argument in which reputations are cheap.

Tautologically, the most harmed people by anything are the most vulnerable. The people most vulnerable to shaming come in a lot of varieties

Materially that is:

1. those who are easily fired or deprived of their livelihood.
2. those who without the money for a legal or PR team.
3. Those from unpopular demographics

But also, emotionally and relevantly to OCD:

4. Those with pre-existing mental health problems or an anxious/depressive disposition and
5. Those with strong consciences.

An inevitable effect of shaming online in which, as one of the commentators above noted, people are treated as disposable cogs, is that it is not necessarily elites who are going to be driven out for their deviancy, but the weak, vulnerable and caring. All in the name of protecting the weak, vulnerable and caring.

One user's comment about how this stuff had gotten under their spiritual armor stood out to me:

“Yep. I struggle with this. For me, it's more internal, though —I'm less worried about the potential effects of being "cancelled" (I do still worry about those as a creative —and also about the actual social interactions that would come with such a thing since I'm autistic and have social anxiety, but those worries are less obsessive somehow) and more about "what if I really am a bad person?" It's sort of a metaphysical contamination theme at it's core for me. I sometimes explain it as the bad parts of Twitter "cancel culture" living in my own head.”

Now if my argument were that we should challenge cancel culture because it can be upsetting to people with OCD. That would be stupid. All sorts of things can be upsetting to people with OCD. My argument is a bit subtler than that. My argument is that the boiling-intense moral economies associated with these online spaces are more dangerous to psychologically and materially vulnerable people than to anyone else. OCD is just one example of this. Other examples include not just mental illnesses, but situations and conditions of precarity in all kinds and varieties.

Tu quoque

One of the reoccurring themes in that thread and others that I found interesting was “why are these people doing this? Don't they know that they've surely done things that they could be canceled for as well?” A while ago I did a survey on justice and mercy. One of the main hypotheses that I wanted to test was that people who had a negative appraisal of their own past would be less likely to engage in judgmental behavior online—or at least express the view that they should be more reluctant to engage in such behavior. Surprisingly I did not find such a link. Yet here was a thread full of people remarking that they think it's foolish to judge when we all have skeletons in our closets. Perhaps there is a link between aversion to judgment and a sense of guilt about one's own past, but only in certain kinds of mental illness? Much to ponder.

Regardless, I think we OCD sufferers have it right here. Judge not, lest when you are inevitably judged, your judges add insult to injury by also calling you a hypocrite.

En passant: “Cancel culture” and religion

At the moment there’s a debate going on about whether that strange iteration of left-liberalism calling itself “social justice” can be considered a form of religion. I think I might write more about this later, but I wanted to make a few remarks.

As an atheist, it’s hard for me to see it from the inside, but religions are often dialectics between mercy and justice. Such tug-of-wars have occurred many times in history, but intriguingly 2000 years ago a radical preacher from Nazareth came into conflict with the religious authorities of his day, in perhaps the most direct iteration of the clash between mercy and justice.

I’m not religious, and this isn’t a religious, blog, but the words of Jesus here have interesting parallels with the present. He argued that in their tendency to emphasize strict adherence to the law over mercy, the existing religious authorities inevitably made hypocrites of themselves, because they preached rules that were too strict and comprehensive for themselves to keep, and preached them without exception. Thus they did in secret what they said not to do in public, corrupting their own virtue, even while they were concerned to demonstrate virtue. He compared them to tombs, bleached white and shining on the outside, but inside filled with rot and corruption.

Genuine concern for other people, he argued, was much, much harder to uphold than a list of laws. It demanded infinite concern and activity, rather than a box checking approach to goodness. It would require being honest about one’s failings over respectability, and seeking forgiveness for those failings, even as we granted forgiveness to others. Genuine concern though, was more authentic than legalism. This is to say, in the dialectic of justice and mercy he championed mercy. In championing authenticity and direct concern over rules and habits, he was perhaps also the first existentialist.

2000 years later as we talk with each other about justice while trying to throttle each other I find that interesting. What would a Social Mercy Warrior look like?

Results of the survey on Dark Secrets, part one

A little while ago via this blog, I released a survey in which I asked people to answer some questions relating to dark secrets and their experience of dark secrets. You can still take the survey here if you like, though I'll be analyzing in this the first 150 responses, future analysis might include any new responses.

I'm going to be releasing the results of this survey in two tranches. This is the first tranche and will consist in results and analysis of numbers and proportions of answers to questions considered on their own. The next tranche will consist of the analysis of correlations, odds ratios, and other relationships between answers to different questions. This is because I'm going to have to do a lot of coding to get the data set up properly for the latter kind of analysis. [Edit: life caught up with me, I was never able to do the second tranche.]

1. Number of responses

The first interesting result to report is that it appears a large number of people, proportional to the number of people who saw the survey shared somewhere, were willing to fill it out. I have experience collecting responses on surveys, and it wasn't notably difficult to get people to respond to this one- the ratio of shares to responses was about the same, if anything it might have been a little better. Given how unpleasant a topic this is to think about, and given understandable concerns about results leaking, this might be a surprise to some. I suspect that the inherent interest in the topic counterbalanced these barriers. Speculating a little, I think that keeping secrets is a lonely and tiresome business, many desire to let their guard down- if only a little- and these sorts of people were likely attracted to this survey.

2. How common are dark secrets?

I defined a dark secret as follows:

“By "dark secret" in the context of this survey I mean something special. I mean a secret about you, something you did or failed to do, which, if revealed, could severely harm your reputation with your peers. A dark secret doesn't have to be objectively bad, it just has to be something which, in your view, would harm your reputation given the context of your life. Homosexuality and gender non-conformity aren't wrong (at least in the survey author's view) but for some people would count as a dark secret. Do you understand?”

Only 7% of respondents said that they didn't have a dark secret. An additional 14% said they probably didn't have one. 24% said that they probably did have a dark secret but weren't sure while 58% unequivocally said they did. Granted, this has got to be due, at least partially, to selection effects on who chose to fill out the survey. Still, I suspect these results probably are at least very broadly reflective of numbers from the general community of the educated, online set who congregate at places where I shared this survey, like the Slate Star Codex Subreddit, this blog, and the sample size Subreddit. The extent to which they are reflective of the general population is something I won't hazard a guess on. I suspect the substantial majority of adults do have dark secrets, though that's based more on a general impression than the results of this survey.

3.1. Severity of secrets- How serious do people think their own dark secrets are in comparison to everyone else's

A result that grabbed my eye was that only about a quarter of people thought that their dark secrets were more serious than other people's dark secrets. By comparison, about half of respondents thought that their secrets were less serious than those of other people. About a quarter thought their secrets were roughly as grievous as average.

The most obvious interpretation of these results is that people underestimate how bad their own dark secrets are or overestimate how grievous the dark secrets of other people are, or both. Another interpretation is that our respondents are correct because the dark secrets of respondents to this survey were on average less grievous than the dark secrets of the whole population.

Another possible interpretation. Assume that the severity of dark secrets has a long tail and is right-skewed. If people interpreted this question as being about their dark secret's severity versus the mean average severity of dark secrets (though maybe not in such explicit mathematical terms), then this result could make sense. A substantial majority could be below average (that is below the mean, not below the median). This is true even if we assume that the sample is representative of the population as a whole.

A promising avenue for future research is investigating whether people underestimate the relative gravity of their dark secrets. For example, we could investigate this by having people share their secrets with the researchers, getting participants to rate the severity of their own secrets, and then having other participants rate each other's secrets (names withheld, of course).

3.2. Severity of secrets- How serious are these dark secrets in an absolute sense?

How serious are the secrets we're talking about here? I think in this regard the most revealing figure is that 61% of respondents thought they had a secret that, if revealed, would cause at least some of their friends to reduce contact with them. 34% of people thought that they had a secret bad enough that it would cost them at least half of their friendships. 12%, nearly one-eighth of respondents, felt that they had secrets so bad that if they were revealed they would have few, if any, friends left.

Now it's very possible that respondents are overestimating the gravity of their dark secrets, or how quick their friends would be to abandon them. It's also very possible that some respondents are underestimating how seriously their dark secrets would be received. We also don't know how representative our sample is. If, however, one-third of the population - or at least of the population you are likely to meet online- is carrying around a secret that would indeed cause half their friends to abandon them, that is an extraordinary and sad thing.

3.3. Severity of secrets- Distress

I was heartened to see that only a little over 10% of respondents reported that their dark secrets caused them severe or very severe distress. Another 16% reported moderate distress while 49% reported slight distress. In a survey filled with troubling results, this was not as bad as I thought it would be.

Still, in the scheme of things, 10% of people experiencing severe or very severe distress is actually a lot of distress.

Am I wrong to feel sorry for these guilt sufferers? Perhaps I should be focusing my sympathy on their victims. Perhaps I should actually want people to be experiencing more guilt. But there is something peculiarly useless about guilt as an emotion. I'm always reminded of something I saw on a Magic the Gathering Card once:

Those who most feel guilt don't need to, while those who most need to feel guilt never do.

Hence guilt is a peculiarly useless emotion in practice. A short sharp stab of it can have its uses, but the long guilt of years seems to lead to nowhere good.

4. The assessed probability of secrets coming out is mostly lowish, but not that low

We asked participants the following question:

“Think of the dark secret you are most afraid of coming out. What do you think the chances are that one day it will come out? Answer as a probability between zero and one, like 0.7, 0.35, etc. If you don't have a dark secret, don't answer this question.”

The median answer was 10%. The 75th percentile thought there was a one-quarter chance of their secret coming out. About 80% of people were in the interval between a 1% chance of their secret coming out and a 30% chance. On the whole, then, people thought it was a real, but fairly unlikely, possibility.

I discarded text answers to this question because they were difficult to interpret. However we had several “it already did” and “in a sense, it kinda already did” type responses.

5. Willingness to pay to protect one's reputation

Participants were asked

“Suppose you were offered a deal by a genie. You could give a certain number of years off your life, and in exchange, your reputation would be guaranteed forever- against damage from dark secrets and everything else from your past. Transgressions in the future could still harm your reputation. What is the maximum number of years you'd be willing to give? Fractions, e.g. a quarter of a year, are acceptable.”

Years were chosen as the currency because A) It is constant from country to country B) The number of years people have to spare- especially among people who filled out our

survey who I assume tend to be youngish- varies less than income does. One survey taker's year is much closer to being interchangeable with another survey taker's year, on average, than a survey taker's dollar is interchangeable with another survey taker's dollar. In hindsight though, years have another problem- it is taboo in our society to trade life for something else.

The median amount of time people was willing to give was nothing. About the 75th percentile was the borderline between 1 and 2 years. The mean, not including the two people who they said they would give all their years, was about 3 years.

20 out of 141 who answered the question said they would give at least five years. Two people said however many years they had, they would give all of them. The overall picture here is huge variance with a big rightward skew. It will be interesting to analyze this in relation to other variables.

6. Horizontal sympathy

Forty percent of people stated that their own dark secrets made them more sympathetic to others who are outed as having dark secrets. This question was of great interest to me. I have a form of OCD that often makes me think, wrongly, that I have severe dark secrets, and makes me worry disproportionately about those dark secrets I do have. I have observed that this form of OCD has made me more sympathetic to people who actually do have dark secrets and get found out. I was curious to see whether this was other people's experience, and in a way, cheered that it was.

7. The content of dark secrets

What about the content of people's dark secrets? Here's some information on the responses we got, keep in mind that response categories were non-exclusive- you could check as many as apply. There was a short custom answer option, but I will not be reporting on any of those results due to the risk of identification.

A lot of them were sexual. The most popular category was “sexual misbehavior not elsewhere classified” with 44 responses. I was disappointed at how popular this category was because I thought that I had set up enough sexual subcategories that we would be able to classify the majority of sexual secrets. Categories in the survey related to sexuality included consensual homosexual activity (5 respondents), cheating on your partner (14 respondents), sexual fetishes between consenting adults (22 responses), incest, whether consensual or not (7 responses) and “Breaches of sexual consent broadly construed including but not limited to harassment, assault, stalking, statutory rape, rape etc.” (19 responses). If you have any theories on what kind of acts might be going into sexual misbehavior not elsewhere classified I’d like to hear them. I can think of a few things that would fit, but nothing that would explain why it was so overwhelmingly popular. I suspect that many responses perhaps did belong to another category, but people didn’t want to admit it to themselves.

Secrets about gender non-conformity (5) were a fair bit rarer than sexual secrets.

Prejudice of various forms was relatively popular- eight responses for racism, ten responses for sexism, and seven responses for other forms of prejudice.

Miscellaneous “destructive behavior” including physical violence (10), taking something you shouldn’t have (13), and destroying an item you shouldn’t have (9) were common.

One of the most popular categories was “Betrayal of trust” with 28 responses.

Family mistreatment other than your own children or partner (3), mistreatment of a partner (4), and mistreatment of your own children (3) were rare. I have a sense that a lot of people may be blind to their own abusive behavior, so I suspect the real figures are higher.

Incompetence (13) and negligence (14) were also quite common.

8. Gender demographics

One thing that will limit the generalisability of our findings- at least beyond these kinds of internet spaces- is the small number of women who responded- only about 17%.

Prefer not to say and other made up 8% of respondents. The rest were men.

9. A bimodal distribution of happiness

Something that was very interesting to me, although it is not directly related to the subject at hand, is that we found a bimodal distribution of happiness, with one centre at 3 and another at 7 (happily the centre at 7 was somewhat larger than the one at 3. What gives? Well, it could be a feature of the population we sampled, but I do not think so. I think it was a feature of the unusual question I was trialing:

“On a scale from 0 to 10, where 0 is the saddest person you know, and 10 is the happiest person you know, how happy are you?”

My theory is that people looked at this, and tended to immediately have one of two reactions “I’m not the saddest person I know, but I’m sadder than most” or “I’m not the happiest person I know, but I’m happier than most”. These respectively corresponded to a centre of gravity of 3 or 7. Why there weren’t more people who immediately thought “I’m about typical for the people I know” is a mystery to me.

10.1. Concluding thoughts- My suspicions about dark secrets in the population

I first became interested in dark secrets, beyond the context of my own strange OCD obsessions, when about two years after metoo I started counting and I realized that I personally knew well over thirty people who had been accused of intimate partner

abuse, sexual harassment, or sexual assault, with varying degrees of severity. Now it must be said that I know a lot of people, and the people I know are of a set who are unusually likely to use (sometimes dueling) public accusations to try and punish perceived offenses. Nevertheless, that's a lot of dark secrets.

My strong suspicion is that a majority of people have, at some point in their life, done something really bad. We'll operationalize "really bad" vaguely as "at least as bad, in the public view, as having had an affair while married". Doubtless, many people don't recognize their own misdeeds. Their brain won't let them clearly perceive what they've done. Others know it, but wouldn't admit it, even in an anonymous survey. But I would confidently guess that at least fifty percent of adults have, at some point, done something "really bad" on this definition (again- greater than or equal to having an affair while married).

But having an affair is only the minimum, many people have done much worse. Going back to my initial point in this section, consider how common perpetrating sexual and domestic violence must be. 20% of women report having been raped, or subject to attempted rape. One in four women in the US report being the victim of severe physical violence by an intimate partner. Many, many more, probably solid a majority, have been subject to some form of sexual violence or domestic abuse, including groping, stalking, emotional abuse, and harassment. Indeed if we use a broad but not exceedingly broad definition of sexual and domestic violence, we'll probably find that most men have been subject to sexual or domestic violence too. I know I have. Now even granting that most perpetrators are repeat offenders, that means there's got to be a fair chunk of offenders out there. As I mentioned earlier, 19 respondents to our survey admitted to "breaches of sexual consent broadly construed including sexual harassment, sexual assault, stalking, rape or statutory rape". How many more have done it, but didn't want to admit to it in a survey, or haven't acknowledged the nature of their own actions to themselves? I'm sure there are graduations of severity here, and probably most of those 19 were towards the low end, but we're still talking some pretty dark secrets here.

What else is there that's seen as at least as severe as having an affair that can be kept as a secret? Lots of things I suppose. Extreme cruelty to animals, gross betrayal of friendship, destruction of a person's reputation in secret, embezzlement, and racist abuse. Even keeping a dark secret can be a dark secret- how many people out there know of a terrible crime yet do not report it and are eaten up inside by that?

10.2-Concluding thoughts- how moral differences can conceal dark secrets even from their bearers

Sometimes people don't even know that they have dark secrets due to contextual moral differences. There are many examples of this but my favorite is Grindr dick pics. Recently the United Kingdom has decided to outlaw the practice of sending dick pics. A (male) journalist in response to this tweeted that he'd been sent several hundred naked photos that he hadn't asked for, and that while it could be irksome, he really didn't think all those people should go to prison. A bunch of people wrote back to him calling an egotistical liar, saying no man would get that many, and he was clearly lying probably to defend his own practice of sending dick pics. The dullards abusing him either didn't realize he was gay or didn't know that this is completely standard for gay men. I can confirm this, I've received hundreds of unsolicited dick pics myself.

Now in the minds of the furious people, what these guys have done in sending me and this journo dick pics is unconscionable. A half step below flashing someone on the street, if that. But the interesting thing about the feelings of these "cyber-flashers" is not that they disagree about this it's that they're not even aware anything they're doing would be controversial. To them, it's just a commonplace, not worth remarking on. Indeed to remark on it is a bit weird. I once had a guy ask me for pics and I wrote back "does that include dick pics?" he wrote back with something to the effect of an eye-rolling emoji and "duh, of course". He found it strange that I would need to clarify that he wanted dick pics- why not just send them straight up when asked for pics? All

these Grindr denizens holding a secret that could get them in trouble if it were publicly revealed, and not even known. For them, it's just natural, a courtesy like showing a prospective buyer a sample of wares.

But the difference between moral contexts cuts deep in a really interesting way. I bet a lot of these very same men if they read an article in the paper about some social media sex pest sending women dick pics would tut-tut and wouldn't even have the thought that there was an inconsistency there. And in a way, they'd be right, because there is a huge difference in the meaning of that act between those contexts, but it's not necessarily a distinction the larger public would acknowledge or accept if these Grindr users' past behavior became public. In the UK for example, it's not as if this cyber-flashers bill has a Grindr exception.

10.3.- Concluding thoughts- Weird darkness

I bet a lot of dark secrets are just weird - random acts of evil are more common than we sometimes realize. I remember I was walking down the street once when I saw a woman on a hen's night start kicking a homeless person and laughing maniacally. I chased after her but I wasn't fast enough to catch her. The homeless guy said he had no idea why she'd done it.

10.4.- Concluding thoughts- why I couldn't ask the most interesting question of them all

I'm sad that I couldn't bring myself to include a question in which I just outright asked people to describe what they'd done. My reason for not doing so? It's possible that, in the future, AI-based stylistic analysis of writing, combined perhaps with a few other clues, will be able to identify individuals based on passages they've written. Moreover, people tend to give away clues about their identity in their writing, even when they don't mean to. Thus I could not justify it to myself from an ethical perspective.

10.5.- Concluding thoughts- Luck in who faces consequences, and moral luck in general

We do and see these terrible things, and yet only a minority of us will face genuine consequences, and the distribution of consequences seems random- at least from the point of view of justice. It sometimes seems to me that we think- at least unconsciously- that so long as you don't get caught for something it's not so bad. Sometimes we even codify it. I once saw a woman express the view that if you married a "real alpha" he was bound to cheat, but if he truly loved you he'd make sure you never found out. Granted, sometimes getting caught can be a signal of special recklessness, or that you're more likely to have caught because you did it many times, but our habit of harshly socially punishing people for things that are relatively common, but are just rarely found, doesn't seem right to me. Still, what's the alternative? There's some really awful shit going on out there, and if we can't consistently catch the people who do it, arguably that's a reason to amp up the punishments so the limited number we do catch can be used as strong deterrents.

Obiter dicta, all of this puts me in mind of the conceit of moral luck. You run over and kill a pedestrian while speeding at 25%+ plus the speed limit? That's a huge deal and is thought to reflect very badly upon you- a time in jail is probably expected, even for a first offense, at least in the USA. Even if you don't go to jail, you're expected to feel bad about it every day for the rest of your life. You go 25%+ the speed limit and don't run over anyone? Ahh man, we all make mistakes, don't worry about it bud, the speed limit should be higher anyway. It's certainly not a dark secret

10.6- Concluding thoughts- Is doing the wrong thing something you can just fall into?

There's a character in a video game I've played -Wrath of the Righteous- called Regill. In one of the most quoted lines of the game, he says that he's sick of hearing excuses:

“Every betrayer has their own sob story to excuse their actions. And each one thinks they are different from the rest, that they alone should be understood and shown mercy. I'm sick of hearing it.”

At the end of the day though, Regill would maintain it's pretty easy not to throw your lot in with cosmic evil. When I was talking about this survey with a work colleague, she expressed pretty much the same point of view about people who cheat on spouses they claim to love: it's easy not to cheat. In the process of cheating, there are dozens of points at which you could stop yourself. Forget this forgiveness crap, this isn't an accident.

We could say the same thing about most awful, reputation-destroying acts I guess. There's the odd exception- the Grindr case we discussed, or Aziz Ansari types who (without any comment by me on Ansari's particular case which I have not thoroughly investigated) are thought to have made bona fide mistakes and paid for them. Doubtless, there are a few other cases at least bordering the thicket of moral uncertainty. For the most part though, to do something seen as really bad, you've got to make a deliberate choice.

Yet though this seems to be true a priori, in practice the species keeps making mistakes and I do not think it's because we are mustache-twirling villains. I think bad people are rare, but people who have done really bad things are not. How are these reconciled? I don't know.

As you look around you and quantify the probable sins of your fellows and yourself how should you feel? When I first started thinking about this years ago, I framed it as a kind of free existential choice. There's no level of sin at which, logically, you must stop loving someone, so that leads you to a kind of unbounded decision, a kind of ethical aesthetic. Will you be on the side of humanity, warts and all, or not? Is there a way to keep loving a species riddled with dark secrets without excusing them?

Poems

Deadwater

I

I recall in tranquillity

Fever-dive hours.

Once I saw a sailboat listing Upon a great-waved sea

The sea was I and so was the boat I could not see any stars

For the blasts of ocean-spray

In what quiet cove can I go hiding from a storm

Blasting up the cartoid artery and flooding through

The cognitive estuaries, over-spilling memory's tributaries?

Tell me where I might make my stand against my wrath?

Might a clever present play the future off against the past?

Am I to live only in the lacunae between foretelling & recollection

In the times between guilt and dread when, exhausted of mental flight,

Whether backwards or forwards, the I drifts in easy content?

We shall build a tower

let us make us a name, lest we be scattered abroad upon the face of the

whole earth

II

Behold, a shattered glass bowl that held doubts

They multiply in shattering

As each beam of light

Crosses every glass splinter

It breeds a new splinter

And a new lance of light

Fecund heresiarch

Absolute clarity lies within

That lit glass rubble but the trouble

Is that so does everything else

As in Borges' library up in that tower

III

Do you know where your right hand is? Walking through a shop and not knowing whether you've assaulted someone heedlessly. Analysing each moment of your past like a sicko prosecutor. The fears iterate by sinister Darwinism, seeking cognitive blind-spots. Did I mutter threats of violence to that child? Did I insult that shop attendant? Mixed memory and aversion form a rancid bin-juice born decaying.

IV

I came to the stairs

There was a wobble in her voice
By each step her voice rose higher
So I rise to her and she calls with greater urgency
And I rise to her with greater urgency
She and I can only meet after escalation shatters
Past the horizon of panic and further-
Past the sea rock of worn defeat
She and I must be one.
I sprint.

V

Imagine that someone came to you in the middle of the night, stepped into your mouth and began to grow through your capillaries. They were not content merely with habitation, their constant insistence was that you must keep grafting dead organs and limbs onto yourself. You become a born-again Frankenstein (don't be a pedant) with all the zeal of a convert to an undead lifestyle. The new limbs are heavy, and stink, and burn up your flesh with septicemic fire and puss-flood, but the man who stepped inside your mouth begs you stitch on more.

VI

The inside of a head becomes lonely as it becomes crowded
The only things that elbowed through those crowds

Were other hauntings

Brief dune-sedge love in salted ground

Warring wrath against money made world

Twin engines of raging-love and loving-rage

Racing for diversion and the exaltation of rebellious motion

Circulation round the track kept my blood in motion

Rammed down winds to bellow my lungs

Political contention, war, courtship, frenetic study

Vain dreams of greatness, discontent

Which gave me a little contentedness

To declare permanent war or endless love

And so to terminate surrender in unutterable resolution “Optimism of the will!”—clenched hands, though they wobble

In the obsidian lands where resistance gave no comfort

Resistance still gave sustenance

Just as all the previous Sugatas

VII

Life is so long. Are you so innocent? You are tired. You dream of a gentle place. You saw it as anyone might imagine it—holy light on wild-flowers, easy with its comforts, free

with its joys. To be such a place it had to be distant from this world and sealed against you.

VIII

Maybe I just wasn't fucking often enough?

Victorian life is better novelised than lived

Hysterical, neurotic, guilty, phantasmal

Maybe I wasn't drinking enough?

A friend called me the Ayatollah

In respect of my beard and sobriety

Hume and the Buddhist sages pronounced that persons are aggregates

without greater unity. I find myself a bundle but there is no liberation here. The parts rub against each other like cans in a grocery bag bruise fruit. Or perhaps I am the curate's egg.

IX

Give me a seabird's wings

On the cliffs, about forty meters over the crab pools I dream of ascending with the gulls,
but higher

Diving and again rising in alliance with wind

What waves perturb the gull are brief

And if it is to end by hawk, that too is brief

Yet I would rise higher still, till I sat on a perch

Overlooking time and the jolting succession of moments

Above the waves of kings, ministers, exchequers

Yet if I am not to reach that exalted perch

I will be low enough to observe the bright net

Of refracted sun that plays upon the hills of water

Give me a seabird's wings

X

Easier perhaps to talk of the accoutrements of terror and the reflections it invoked.

Easier to do that than to photograph medusa. Yet I do remember being confused as to whether I was more guilty or more afraid. It seemed important that I be more guilty than be afraid, but it is hard to feel guilt while facing knives. Consequently, I felt supplementary guilt at my thin guilt.

We shall build a tower

let us make us a name, lest we be scattered abroad upon the face of the whole earth

XI

The future is boundless, not only ahead but sideways
The patterns of your inferences only ever ape
The subtle causal chains which bind the forward momentum
Of the world whose surface you cling to
The mind is stretched between times and possibilities,
Beyond any accommodation by mental sinew and bone
The heart successively roars and fizzles

XII

I came to the living room
And it was filled with ash
Though I never smoked
Or sat by fire
I made an ink of that ash
And began to write these verses upon my arm

XIII

He is there, and I smile into his oblivion
He never loved you, so ideas of romance
Had the character of Banach-Tarski's sphere
He is gone now, other suburbs, other worlds
I do not miss him, except on special occasions

My affections were never lost, except perhaps at the first moment

Dead on arrival

Yet still worthwhile

It is right to rebel against most things

But not you, oh sweet tyrant

It's good odds you kept me breathing

IXV

We do not sit upon heaven's throne

Nor are we the rebel, cast down like a slash of lightning

We are the flesh that raised our gaze

Half wondering, half begging

The dance is ending, where is the bridegroom?

XV

How rash are those who clamour for justice? (I have been among them)

Life is wide, deep and changing.

We are excesses

Of identity, act, motivation.

Of miscalibrated judgement and selfish grasping.

Do you think you would be clean under heaven's eye?

Were there a book that contained each numbered thought and small deed

Of yours wouldn't you shred it, burn it and eat the ashes?

I wouldn't. I would give you that book. Press you to read it.

I do not think you would like me, but my terror is to be misunderstood

I fear that you will think I am a different kind of monster than that I am.

So I give you my promise, that should an angel scribe that book

I'll give you a copy.

And I promise that if you ever give me a copy of your celestial biography

I'll try to shut the my eye of judgement and open that of mercy

It's simple self-interest. Chesed pro chesed.

XVI

Can we remember pain? In our mind's eye we might

See rose fluids or, under that, a startling glimpse of pearly white

Laid open by a scalpel. We shudder back. We peer forward.

But who has the pen by which to bind agony? "Sharp", "dull", "throbbing", "irritating",
"intense"

Wholly feeble, as if a snake tried to wander with its vestigial leg bones

But that is where we find ourselves—thirsty for conveyance in a desert of names

We can only hope to articulate pain through our inarticulateness

Just as, by chance, static on a television set captures a snowstorm

I remember wandering the streets, sobbing and calling for divine fire to kill me and all

the other wicked. As I wept I listened to pop on half smashed headphones. What would it take to make you march through city streets weeping and calling the fires of an unknown God?

XVII

I ascended to the attic

To store, retrieve, invent

A mnemonic parade

Without volition my hands

Raise the dust in small incantations

How does one dislodge a fake memory?

Or terminate the routine of shuddering

I see

He and she are here, interlocked eye-beams

I am not in either eye

In this attic I lay in the pattern of my veins I am sinews.

Whether these gobbets

Be thought or flesh I am in neitherway free I am chained by my own substance.

Above me powers contend in the air.

XVIII

Think now

Life has many cunning passages, contrived corridors And issues, deceives with

whispering trepidations, Guides us by vanities.

After such knowledge what forgiveness? Forgiveness after such knowledge what? What forgiveness after such knowledge? Knowledge what forgiveness after such? Such knowledge what forgiveness after?

IXX

In metamorphosis the tissue is not merely subtracted from and added to inside the pupae, rather the whole flesh devours itself, save for microscopic clusters (imaginal bodies), becoming a soup of cells. What unites both life-stages is scarcely more than a double-helixed teleos. Yet memory persists.

We shall build a tower

let us make us a name, lest we be scattered abroad upon the face of the whole earth

XX

If I could but seize the wax of Icarus

The tailor of Ulm's fabrics

Etana or Bladud's crown of feathers

If I could but fly, I could seize the sun's silver

Forge a mirror by which to demonstrate

The storm that rends the head

Of some shivering soul you know

Forgive a thief that stole for you and
Shelter all, for you, cannot see their weather

XXI

To find a point of collapse at which loss and victory die.

And that sea is now

A vast lake that

Night or day

Forms a perfect twin

To the sky

Over the stones of the tower

Drift currents and sweet, lazy fish

The waves will dance again

But I might hope to dance

With them

Afterword to Deadwater

A word on credit. This poem is allusive to the point of plagiarism, and past that (about 5% is lifted from other poems). My purpose is to convey an experience with all that I have and I'll gladly steal words for that. I have no concern to prove myself as a poet, only to tell the story as well as it can be told.

The debt to T.S. Eliot is obvious, even in the title. The debt to the Aiken's Tetelestai and the Romantics (including Eliot perversely read as a romantic) is less obvious. It's very much a poem about me, and I apologize for that vanity. My story is not unique. My particular kind of OCD based on a fear of harming others is quite common. Yet few talk about it for fear of seeming like a dangerous weirdo. It is an inherently self-concealing form of mental illness. Especially as I've gotten older, I've tried to avoid the narcissism of self-display even in an anonymous form, but I want to show you this story, lest it be scattered everywhere among the nameless like me, and forgotten.

Upon seeing a rose encased in ice

The night air
Throws a glass cloak
For modesty
Around the red
Offering
Offering a little pause
In time from withering
Pleading no end to
That redness

But the sun reclaims for a jealous sky
That water again
Soon after, rotting petals fall
I am sorry water could not hold time

Untitled

Do I remember the pyroclastic light?
The rhythm sun of the sex discotheque
Robbed in warm water and silky skin
The limbs slide over silk in soapy suds
The need to own and be owned, fused
Clothed in bridal white and rapid waters
Sequined in bubble bright red gleaming light

Dying in the rain

Do you think there was blood in
The mud in that flood as the
Rain cascades down down down down
As the wicked ram against
The wooden gates of the great
Arc stark upon the green hill?

Or do you think the wicked
Beheld their children and saw
Their mothers and their fathers
Saw then that though they were bad
They were people cold in the rain
And how merits man to drown?

Did Noah floating on the boat
Remember their names, their sins
Or did he just see their faces
Sad, scared in that first rain
No man nor woman so bad
That they were not dying flesh

Narrow escape

I lost myself in the heaven of stars
Circling the drain of being, of ending
Your cheeks like the slopes of deadly mountains
You are direct, sharp, shapely, kingly
I wanted, right there, to hold you snugly

Your hair tumbled like a silken crown
Your red towel loose as if a regal gown
Your voice drifts, laden in incense of power

We talked about the best song by tenacious D
We talked about trouble composing harmony
I told you I found you once on Grindr
Thought you were catfishing when you caught me
Now though, here you are, bereft of catty whiskers
You said you were an actor, I wondered
What you'd made and whether I'd ever see it

I wish I'd asked your name, your name though
I am glad I didn't ask your name though
Your ninety seconds of speech haunt me still
Another ninety and I'd have loved you.

Glide your gaze across blank walls

Glide your gaze across blank walls
Seeking for purchase on those
Perfectly smooth rocks, illegible
To all interpretation or penetration

Consider a dimple in the wall
A line or knot in the wood making
Earthbound constellations a practice
As old as buildings- older indeed
Since there have always been caves

The school child and the worker alike
Achieve a state of agitated tranquility
A state that is neither inside nor outside
Time, or rather time becomes an endless
Field- again, without purchase, again in
Repetition, tick, repetition, tick, repetition
Dull, dull. Dull, dull. Dull, dull. Dull, dull.

Such strange illuminations on the mental retina
Like a lantern made to cast a special shadow
Can generally only be achieved with illicit substances
But in the schoolyard, the office, the factory
Eternity is free. Churlish mind, why do you
Not grant such in the brief hour with friends
In the embarrassingly short minutes of fucking
In the days with dying mothers, or even
The last years of an old dog's life
You have made eternity unequal in each hour

Until we die, and from that moment on, all divisions
Upon time receive their just allocation of oblivion.

Dew

Dew flies in
the sunlight

Conversation in a garden about a trampling

A garden. Trees are swinging. Stars are spinning cold.

That cold up above parade, liminal lights.

The dreadful lust of roses beneath moon shade.

...of course, he speaks

“I trust not justice elephants dispense, hence

Though that there bull trample twice I trust

Him not once and wish the woman more mercy

In next carnation, so she be free of tusks”

So bad- I'd like to fuck him till he's mute

...Impossible creature, he owns all the rules

...It's his eyes the stars spin in

...I love his stupid genius

...I'll be here so he can discourse all night

...I hold him from behind

...Cupping him against astral and boreal winds

Futility

New shoots

On a felled trunk

Moonrise

The night before moonrise is old
The weary birds conclude their skirmishes
Singing out agreed partitions till tomorrow

But the night after moonrise isn't young or old
No time (or all) passes under the lunar-lit stars
Eternity strokes the earth and cicadas take up
The star-chorus

Untitled (slam poetry entry)

I recall an evening, out of my mind
Prone, belly skyward, dissolute below the stars
And wondering
At my absences- at my trite completions
But on that evening I was distant from my own heart
So I beheld distantly the gorges, pinnacles and crevasses
Of myself
And in my relaxation I permitted
Wonder at my own hesitations, fractures, fearings
Seeing my being in a kindly, warm and patronizing light
I have always wanted to return to that place
Where I ceased to be yet I was not dead
Where I gazed upon my own living soul

And I have heard it said by the poet
That with each dart and turn
A sparrow annihilates a world
Enters a new one
But in the lidded cauldron of the human skull
Steam has no such exit

We become wise
In the memory of gathered thoughts
Carrying them, like aching joints
Till we are arthritic in the mind
Hauling the cold weight
Of thought ghosts - Of memory ghastrs
Homesick for a distant place,
Soft in mercy

By a lake with friends

We are sitting in a wild garden
On a warm twilight, with cheap wine
There are about twelve of us
We laugh over and over like waves

And I know I will always be here,
Twilight won't end, I have my peace
By a lake with friends I lie, I laugh
By a lake with friends I lay, I laughed
Even now, it is so that I lay, I laughed

Break silence

You don't
Have to break
Silence

You can just
Walkthrough

The Ballad of Tim

In the distance at the fork of a river
Off the coast of Canada, a grizzly stands
He does not need a mysterious forgiver
He is too simple to do a thing wrong

Long have I longed to be known for
Faults and virtues. To sing of myself and
Leave none in doubt of veracity as I tore
A copy from the account book of the soul

To have a disease which doubts and condemns
To create a mess in the heart so dense
You wish to open the door to show your friends
Air it, get help carrying it all outside

And if showing the truth made them hate you
That would be a fine judgement and an end
You'd hope to love them still no less than due
Sunsets over the windswept cliffs you love. Be here.

It's morning and the cosmic hand has swept
The cold dark sky and its starlight guards away
Even in the secret chambers of sleep I wept
And the sun won't change that, yet it is warm

I dream of a paradise that I cannot reach
For I am human and would unmake paradise
At some unimpeachable place they come to teach
Each of the seven bright arts of the gentle-kind

In that place there are flowers and holy lights
Because I want to be there I cannot
And for that yellow-warm light to glow
That place was sealed against all like me

Let me be gentle so when the world is rough
And I pray, heaven won't see a hypocrite
And I won't wear the crown of proud white bluff
And if I can't be gentle, let me not pretend

As my flesh rots my mind lashes out weaving
In the void projects, truths and lies to sustain
A life after life, in terror I am heaving
Death is already within me so I run, write, fight

I found you searching for communists and artists
Cyber nymphs led me to where you sat
Algorithmic woods by river of hope

Ibis, peacock, hen and falcon hawk and dove
Skylark, raven, kiwi, albatross and hummingbird
I wish I could open the aviary of my love
But you appraised me and found in me light

We are dying all the time. More or less frantic.
Hurried or unhurried but No less dying still.
We are unbeing, decaying, necromantic
Please hold me till I pass on from this world

I miss you - a lyric

For someone I have never met
Strange thought, I must confess
I miss you, oooh

I miss you in summer moonlight
I miss you in winter sun
I miss you, oooh

One day by pygmalion's chisel
Or one day by explorer's eyeglass
Or one day by a mystic canto spun
I'll find you.

When the branches are too cold
To grasp onto their leaves
I miss you, oooh

When I am scared and naked
When I am soaring unbound
I miss you, ohhh

One day by pygmalion's chisel
Or one day by explorer's eyeglass
Or one day by a mystic canto spun
I'll find you.

When I am veiled in glory
Or draped in humility

I miss you, ohhh

Beneath star, sun or cloud

Above peat, ground or concrete

I miss you, ohh

One day by pygmalion's chisel

Or one day by explorer's eyeglass

Or one day by a mystic canto spun

I'll find you.

On reading your own writing

Reading your old words, at least once you're out past the cringe
Is like reviewing a box full of glass and plastic gems
That you loved when you were a child, ruby or sapphire
What matters is not the authenticity or provenance
What matters is a certain chance interaction with the light
It's not that you see the glint and you're back in childhood
No, you see the glint and suddenly you're gone altogether

Stuff that I was on the fence about including

Why this section?

Few have grasped -in their bones and not just their skull- that there are *no space limits in online publishing*. This is just one of the many ways people haven't adapted to online publishing.

If I'm on the fence about including something, it seems most logical just to chuck it in, and add a disclaimer at the front that it may not be as good as the rest. The reader can then just read it or not read it at their leisure. Why cut it?

Some of the stuff in here reads fine, and is important, but just seems like it won't be of interest to much of my regular audience because it's on a niche topic. Other stuff is (in my evaluation) an important intervention on an important topic, but doesn't quite nail the expression. Some of it, sad to say, is just not as good as the rest of what I have in this volume. It doesn't *quite* make the cut.

Scientific realism and the moral uses of science

The Stanford Encyclopaedia of Philosophy defines realism as:

“Generic Realism:

a, b, and c and so on exist, and the fact that they exist and have properties such as F-ness, G-ness, and H-ness is (apart from mundane empirical dependencies of the sort sometimes encountered in everyday life) independent of anyone’s beliefs, linguistic practices, conceptual schemes, and so on.”

By scientific realism I mean here, roughly, realism about entities posited by science, with realism defined as above.

Here I want to give an argument that anti-realism about science cannot capture some of our important practices. The examples I have in mind are from the social sciences and biological, but I suspect examples could be found entangling other sciences. I will argue that, insofar as you follow these practices, you are implicitly dedicated to realism. Whether you take this as an argument for realism, or whether you think it is simply so much the worse for our ordinary practices will be a matter of epistemic taste.

Hilary Putnam, in an article I have lost, once suggested that the really interesting failure of scientific anti-realism is that it can’t explain our practical reliance on sciences. Speaking about Popper’s falsificationism he suggested that the falsificationist approach of treating theories as interesting ways of thinking that we haven’t yet shown to be false misses an important dimension- our confidence that we can rely on them to keep being at least approximately true in the future, in domains as diverse as medicine and engineering. How can, for example, falsificationist anti-inductivism explain practices like safety testing where there is a clear implication that we think past results give us reason to be confident about the future- something to rely on in a practical sense?

But let's suppose that anti-realists could give an account of why we so often comfortably rely on scientific theories to tell us what will happen. There is another problem- one where we rely on science not so much for its ability to predict observations, but for its ability to tell us how things are because we have a direct ethical interest in how things are in themselves- and not just in predictive success.

Consider the psychometrics of happiness, a booming area of philosophical investigation due to the work of authors like Alexandrova. Suppose that we are deciding whether or not to have school start later or earlier, and we have evidence that teens are generally happier when they sleep later and go to bed later. Anyone who thinks that happiness plays an important role in ethical decision making is going to factor this evidence into their decision-making, but it's not clear that the scientific anti-realist can explain this reliance on social science. Psychometric happiness is, after all, a posited but not directly observable construct. If the relevant science does not give us any reason to actually believe that the children are happier waking up later, then it is hard to see why we should act as if they are happier waking up later- with all the intrinsic value that accrues.

If you want an example from outside the social sciences- consider DNA evidence in forensic science. We really rely on forensic science to tell us whether or not Bob *did it*. That is whether he actually truly did it, not whether it will make certain types of prediction more tractable or anything like that. Mere prediction of future experience, for example, isn't enough, for the practice of forensic science to make sense in terms of (most people's) concepts of justice, we need to have reason to think it actually tells us how things were.

Carving up the philosophical terrain around personal identity a little differently

Many people are aware that there is a debate between the psychological and bodily continuity theories of personal identity over time. I want to carve up the logical landscape in a way which introduces a second, fully independent axis substantive vs pattern continuity. According to the substantive view of personal identity, you survive if whatever object makes you up survives. According to the pattern continuity view, you survive if the pattern that you consist in continues, even if that means destruction of the object that currently instantiates the pattern which makes up you, followed by its replacement with an object that continues that pattern. The substantive view is often conflated with the bodily view, and the pattern view is often conflated with the psychological continuity view, but as we will see, such they are conceptually independent.

In order to demonstrate this, let's consider two classic cases which, when considered jointly, none of the four possible combinations (bodily substantive, psychological substantive, psychological pattern and bodily pattern) gives the same array of answers to:

1. A brain transplant

If your brain is placed in a new body and your old body is destroyed, have you effectively "changed bodies", or have you simply died?

According to both psychological views of identity (pattern & substantive) you survive a brain transplant in which your initial body is discarded. According to both bodily views of identity (pattern & substantive), you do not, since you are constituted by your whole body, not just your brain. So far we are in agreement with the standard account which

treats all psychological theories as pattern theories and all bodily theories as substance theories.

2. Teleportation

If you are annihilated, then reconstructed elsewhere by a teletransporter, have you survived?

Here's where we diverge from the normal account. Teleportation is often thought to separate the bodily and psychological continuity theories of survival, with the psychological view contending that one survives teletransportation and the bodily view contending that one does not. Instead, in our taxonomy, what this case really separates is the pattern and substantive axis of views on personal identity. According to both pattern views of identity (bodily and psychological) you survive teleportation. In both cases the pattern or arrangement that, according to these views, consists in who you are, is continued, since the pattern of both body and mind is recreated. Also in both cases, the substance of what you are (either the actual mental states instantiated in the brain,

Should you care about that issue?

Does it:

1. Involve money or the economy
2. Involve the military or foreign policy
3. Involve criminal law. state violence or mob violence

If the answer to all of the above is no, it's probably a bullshit issue that's being used to distract you. Even if it does involve one of the above it may still be a relatively small issue that shouldn't be prioritised, but at least it's somewhat important.

This is not a plea for crude economism. For example, the right of trans people to access bathrooms is absolutely an issue, because it involves criminal law and violence.

Immigration? Absolutely worth talking about under points 1 & 3. Racist policing? Absolutely worth addressing.

On the other hand, some idiot celebrity saying dumbshit on Twitter? Almost certainly not an issue. Arguments over Starbucks saying happy holidays rather than Merry Christmas? Not an issue. Trump looking like an idiot? Amusing, but not a real issue. Casting for a movie? Not an issue.

There are exceptions. Same-sex marriage was, I think, an important fight to win, even though the connection to any of those criteria is somewhat tenuous. An argument can be made that it involved 1, given the economic rights afforded to married couples, but even if these economic had been fully equivalent I still think the campaign would have been (somewhat) important. Nonetheless, you should start to get suspicious if an issue doesn't tick one or more of those criteria.

Conservation of moral status under misfortune

“His father was a drinker
And his mother cried in bed
Folding John Wayne’s T-Shirt
When the swing-set hit his head”
–John Wayne Gacy Jr, Sufjan Stevens

I want to outline a (somewhat) rigorous way of understanding and defining “mitigating factors” when it comes to moral judgements of praise and blame, without appealing to any view in the debate on free will.

Regarding the serial paedophile, torturer and murderer John Wayne Gacy jr, the line in the above verse: “His father was a drinker” is an understatement. Gacy’s father beat him, sometimes to unconsciousness. He suffered several severe head injuries as a child. These caused him to experience periodic blackouts throughout his adult life.

There is evidence of an association between both child abuse and head injuries on the one hand and serial killing on the other. This study: The incidence of child abuse in serial killers study on sexually motivated serial killers finds that while 68% of them survived child abuse, only 30% of controls did. Serial killers were also six times more likely to have been physically abused -like Gacy was—than controls. This study: Neurodevelopmental and psychosocial risk factors in serial killers and mass murderers reports on widespread head injury among serial killers. It is far from the first to do so. A link between head trauma and other forms of brain injury and serial killing seems to be folk wisdom in the field.

Two studies alone isn’t definitive. But there is at least a prima-facie case for an association between serial murder, and misfortunes like these. Serial killers are a difficult population to study due to their rarity, so one option is to look at violent

offenders in general. One imaging study High prevalence of brain pathology in violent prisoners: a qualitative CT and MRI scan study of violent offenders found they were five times more likely to have sustained a brain injury than non-prisoner controls.

One objection is that because the majority of abuse and head injury survivors don't perform horrific acts, we can't attribute Gacy's actions to these conditions. This argument confuses different kinds of enabling conditions. You couldn't take any random person, add these elements and have a killer—no one is claiming these are sufficient conditions. Rather, there is a very good chance they were part of what set Gacy on his specific path. Different people react to different stressors differently.

So Gacy was quite probably murderous due in part to the misfortunes he suffered. In me this evokes a sense of sympathy for Gacy, even as I remain appalled at what he did, and, to be honest, disgusted by him. Such arguments are well trodden ground. They are a favourite of everyone from philosophers to opinion column writers. They are often made in the context of arguments about free will. Bluntly, you've heard this all before right? Yada yada, tough childhood, yada yada, their behaviour was determined etc. etc.

Here's the point I want to make. I don't think you need to talk about free will at all to think these factors should affect how we see Gacy. In fact I think centring the debate on free-will muddies the waters, making it unnecessarily metaphysically sectarian. The bare fact that Gacy could have been someone different and wasn't in part because of a misfortune that happened to him moves me to pity—no metaphysics around free will needed. A lot of people I've talked to hold similar intuitions. This is important because debates over free-will are intractable. If we judge a certain class of people deserve clemency, then it's best if we can articulate this without appeals to our metaphysical views. Views on controversial questions in metaphysics like free will v determinism are hard to unite the public around, but, perhaps surprisingly, there is often much more common ground around ethics.

I suspect that the underlying ethical intuition is something like this: *Conservation of moral status under misfortune*. Consider the portion of the population who have blamelessly suffered some misfortune X. Call these people Xers and the people who have not suffered X, call them NXers. X hurts you and may cause you to behave badly. An omniscient, rational and morally good observer would regard NXers and Xers as overall morally equal on average, proportionally adjusting the praise and blame due to Xers accordingly.

This seems to me to be a principle of justice. It holds because we should not, in general, blame the unfortunate more than the fortunate, or praise them less- to do so would be to morally favour people whose lives have gone well for them. We should accordingly adjust the “moral standing” of each Xer upwards. In some cases, like that of John Wayne Gacy Jr, this is still not going to be enough to get them out of the “bad” region, but it does make them less blameworthy.

We might even mathematicise it like follows. If a certain misfortune—like a head injury—makes you five times more likely to commit some horrendous act, that act is five times less blameworthy if you do commit it. So we have a (semi) formal basis for understanding mitigating factors. No misfortune should make you, in expectation, a more blameworthy person.

We have at no point referred to determinism vs free-will. The principle is workable—and attractive—even if you believe in the most demanding concepts of free will. This is so long as you accept the empirical premise people who suffer certain kinds of misfortune are more likely to do certain bad things. Of course the principle of the conservation of moral status under misfortune as I outlined it is far from complete and perfect—what if there were a kind of misfortune the suffering of which happened to correlate with

Some stolen ideas on dynamic memetics

You've doubtless seen the succession of ideas in subcultures. The hard left for example went through a "check your privilege" phase, but that phrase would now mostly evoke eye-rolls and is now seen as symptomatic of excessive individualism. The liberals on the other hand have taken it up with enthusiasm.

There's a concept called plant succession. Plant succession suggests that, for example, when an area gets burnt out or clear-felled, and new flora starts growing, there's a natural order in which it tends to happen. Certain plants, so to speak, specialize in particular stages of the process.

Call this rough idea, applied to memes dynamic memetics. It's the idea that there's a process by which ideas replace other ideas. I was thinking about all the explanations from sociology, dialectics, etc. of how to categorize the dynamics of memes, and how one meme follows another- this essay followed. It is a collection of stolen ideas. The terms and concepts I use are taken from areas like sociology, philosophy, biology, etc.

I am not endorsing any of these ideas individually or collectively. My own best guess is that all of these play a role at certain times. No predictions follow from this essay alone, but it gives us a language that could be used to formulate predictions.

I divide these approaches up into two classes. Material factors deal with the lives and circumstances of the people holding and influenced by these memes, and ideal factors which relate to the content of the ideas themselves.

Material factors

1. Functionalist theory- our ideas are selected because they help society work, for example, by holding it together, preventing conflict, and making it resilient.

Dynamically speaking, change happens to meet new social needs.

2. Conflict theory- our ideas are selected because they help certain factions, esp dominant factions. Dynamically speaking, change happens to meet new needs in the contest for social power.

3. Individual interests theory- The selection of ideas is best understood in terms of the interests of individuals. As things change (and that change may be partly driven by ideas) the interests of individuals change, and so their ideas change.

4. Standpoint theory- The selection of ideas represents not so much the interests of societies, individuals or groups as their point of encounter with the world, and the terms on which it happens. This can be tied, for example, to conflict theory (ideas represent the way this or that group in the social struggle encounters the world) or individualist theory. As encounters with the world change, ideas change.

5. Symbolic interactionist/interactionist theory- ideas develop in the engagements between people. As those engagements change, memes change. We could understand this in all sorts of subcategories- for example, ideas develop in the conflicts between individuals (a sort of miniaturized conflict theory). Ideas develop in order to facilitate the interaction of individuals (miniaturized functionalism). Or ideas develop to match the standpoints of people in conversations- their joint encounter with the world and not so much their interests (standpoint theory).

Ideal factors

6. Founder theory- The selection of ideas is biased towards whatever ideas were established first. Emphasizes the relative lack of change, and continuity under superficial change.

7. Reaction theory- Opposite of founder theory. The selection of ideas is biased against whatever was previously in vogue. Just as if you see a lot of red for a while your eyes will be biased towards seeing green (opponent processes theory), so we are biased against whatever was in vogue five years ago.

8. Dialectic theory- The selection of ideas in some sense “evolves” like a chain of reasoning, with new ideas a logical progression on old ideas, in some sense. The exact sense of progression will vary, a common idea is that they react to the weaknesses of the old ideas, but not so much by reacting against those old ideas, but by improving on them or synthesizing the insights of ideas once thought incompatible.

9. Truth theory- ideas gradually approach the truth, or in domains where truth isn't in question (e.g. ethics) a certain comprehensiveness and fullness of perspectives. Individual people may be poorly responsive to evidence, but while the arc of history may be long it bends towards the evidence. This is similar to dialectic theory, with an added realism that is not essential to the dialectic theory.

10. Dirty competition theory. The succession of ideas tends to favor ideas that compete more ruthlessly, by socially or physically attacking competitors, leading their adherents to band together and support each other against rivals, etc. Think of monotheism sweeping away polytheism and variant monotheisms. This might lead us to think that the competition will get fiercer and fiercer over time. However, there are countervailing factors (contact with reality giving selective advantages, the reaction against stifling orthodoxy, etc.)

Mixed/uncategorizable

11. Downstream theory- Certain subcultures are downstream of other subcultures with respect to ideas. The 'advanced' subculture adopts new ideas first, and then the 'delayed' subculture takes them on later- sometimes in a modified form. Sometimes, if one is being cynical, once it has found a way to sanitize them and make them safe.

Money and the sceptic

The other day I saw someone put forward an argument against redistributive taxation based on a study, to which another commenter replied- 'but that study is from the Cato Institute'. Many people jumped in to object that this was an ad hominem argument, and the study should be evaluated on its own merits, Cato institute or no. This ordinary interaction on the internet got me thinking about the social epistemology of such claims and why I think we should not be so quick to dismiss ad hominem arguments in these situations, at least not entirely. Rather we must remain resolutely conscious of the social incentives shaping the discourse we are in.

Consider the following thought experiment:

You're aboard a spaceship and you crash land on a mysterious planet. To your surprise you discover an advanced technological civilization there.

Upon discovering the locals and achieving communications, you are asked to resolve an ancient dispute. You, they feel, are in an excellent position to be impartial. At first you try to decline the offer, but they are very insistent—the issue simply must be resolved one way or another.

They explain what the issue is. The juice of a certain fruit needs to be distributed between them. Ancient custom has acknowledged that for various reasons, some of them are entitled to vastly more juice than others. The issue to be decided is whether a more even distribution of juice should be adopted, or whether the customary distribution should continue.

The juice has the property of allowing its imbiber to go with less sleep than they would otherwise require. These creatures normally sleep for three quarters of the day, but those who possess the most juice can go for days without sleeping. This means that they

can spend more time writing and discussing ideas. Very often the juice-rich write and think about the topic of the optimal distribution of the juice. They most often come to the conclusion that the existing distribution, or a slight tweak on it, is optimal—although a minority of them do support redistributing the juice. Sometimes they give out the juice to intellectuals of their acquaintance who then write about topics including the distribution of the juice. Since these intellectuals were selected by the juice owners, their writings most often conclude the current distribution of the juice should stay as it is.

Considering the situation, you find yourself with a problem. You suspect that the existing literature on the optimal distribution of the juice is probably quite biased, that results which seem to support its unequal distribution are found more often and receive undue prominence due to the structural advantages which support the status quo. However you are not an expert on the juice literature, it is very confusing and contains many mathematical symbols. There are a lot of stats, and every time someone puts forward a stat, someone else says that this is actually a misunderstanding.

What should you do?

We can debate exactly how much weight you should put on the lopsided origins of this society's thinking about the distribution of the juice. We would all agree though, that in the absence of the cognitive resources necessary to sort through the theories and arguments in ideal detail, if one has to make an assessment on the basis of one's limited information and intellectual powers, some weight should be placed on the reality that the side supporting the status quo commands greater resources, and this is bound to make its arguments appear stronger than they otherwise would.

Lifting the thin veil, if you haven't already guessed, the juice is money. The moral of the thought experiment is that in assessing the arguments for and against redistribution you should absolutely be acutely aware that there are not a lot of think-tanks funded by

poor people. Even if you happened to be a trained economist, your mental capacities are still limited, and you do not have infinite time to check all things. You should, at a bare minimum, prioritise listening to and finding material by the side of the argument that is likely to be less well funded since you are less likely to encounter such material organically than you would be if both sides were funded equally.

Postscript: on the liberal media

Since I came out with money and the sceptic, the most common argument I've heard against it is that a majority of think tank and media money is controlled not by the right, but by liberals. The premise is true (or mostly true) but it doesn't negate my argument. When I talk about the pro-centralisation-of-the-juice factions in my thought experiment, I'm talking about a group that includes almost all liberals and almost all conservatives. Thus whether liberals or conservatives get most of the media or think tank funding is beside the point, because from my point of view, on questions of wealth distribution they're basically on the same side. I am counterposing them against a tiny minority in public life who support fair-dinkum, honest to god, wealth redistribution. These people get almost none of the funding.

On the perils of contrasting niceness with kindness

At the moment there is a low level cultural conversation going on about the difference between being nice and being kind. We are reminded that these things are different. It is possible to be a gruff old bastard with a heart of gold. It is possible to be sweet but selfish. I see this a lot on Twitter. Some people are even try to present them as opposites:

[In the original I had an image sourced from the internet purporting to counterpose niceness and kindness. The kind “speak up” whereas the nice “are quiet”. The kind are involved in “healing” whereas the nice are “toxic”. The kind tell the truth, whereas the nice lie to keep the peace. The list continued on for more dichotomies, and the clear claim was Kind=good, Nice=bad]

I am wary of this distinction for a couple of reasons. Let’s start with definitions. By nice, I mean exhibiting a warm and friendly demeanor and being polite. By kind I mean being willing to make sacrifices on behalf of the well-being of others, and refraining from taking advantage of others. Here’s why I don’t think we should be so quick to pry niceness and kindness apart.

The statistical relationship

Open Psychometrics is a website that provides psychological tests people can take. You can opt in to make your results available to researchers, so I looked at the big 5 test they had—specifically the agreeableness facet (“agreeableness” is a composite factor including, among other things, both niceness and kindness). I took the two questions in the agreeableness facet which were obviously related, positively or negatively, to being nice, viz:

“I insult people.”

“I make people feel at ease.”

And added up user scores for each (or rather added the question about making people feel at ease and subtracted the question about insulting people). Making a composite score of niceness.

I then took the three questions that seemed clearly related to being kind—positively and negatively. Subtracting “I feel little concern for others” and “I am not interested in others problems” and adding “I take time out for others”. Making a composite score of kindness.

I then correlated the two composite scores, finding they correlated .37. This is a moderate correlation by the standards of social science, but the correlation is likely attenuated because the correlated scores are based on only a handful of questions—and very difficult questions to judge accurately about oneself regardless. When variables are measured unreliably, this makes the correlation between them weaker. This means the real correlation of the underlying variables, measured in a more reliable way, is doubtless much higher. You can read more about correction for attenuation here if you are interested: [Correction for attenuation](#). I used some lowish estimates of the degree of unreliability and got a corrected score of $r=0.5$. I would bet a lot of money that the real relationship is much stronger still, for various statistical reasons I won't get into here.

So while these variables may be, in theory, different things, in practice they tend to go together quite strongly. There is a robust tendency for the same people who are kind to be nice.

This is an important point that it's worth remembering in the social sciences (everywhere, really) just because two things are conceptually distinct as ideas, doesn't mean they easily come apart in the real world. The idea of having a kidney is very different to the idea of having a heart, yet every single organism with one has the other.

For ease of mental labeling, you might think of this as the “conceptual distinctness fallacy”—the fallacy of thinking that the degree to which concepts are distinct from each other on a conceptual level determines how often the variables they track come apart in the world.

I would note one important caveat here. Be careful in applying this principle cross-culturally, because there are different norms of politeness across cultures. Just because someone comes from a brusque or abrupt culture doesn't mean they're ungenerous. The relationships between variables that exist at the level of individuals aren't always the same as the relationships that exist between variables at the level of groups.

So yes, there is a relationship between kindness and niceness, and that leads me to a further worry. I am concerned that people use this idea of “niceness isn't kindness” as a crutch to excuse their own verbal cruelty to others. It can become a way to avoid difficult questions about how we treat other people. “Oh sure I told that person on Twitter to eat shit, so I'm not nice, but don't you know I'm kind and that's way more important”(1). People can use this line of reasoning to avoid self-improvement. It could even be that if niceness and kindness are related on a deep level, letting our niceness atrophy may reduce our kindness as well.

But moreover, the effects of not being nice and being unkind are more similar than people realize—they can both fuck people up. Humans are fundamentally social creatures we are intrinsically—without our voluntary choice—effected by the opinions of others. Being insulted hurts. Being shunned or ostracized is extremely hurtful according to a wealth of psychological studies. One of the main, consistent results from the literature on ostracism is that being ostracized hurts a lot more than most people realize. See for example: “The Ordinal Effects of Ostracism: A Meta-Analysis of 120 Cyberball Studies” and these effects are particularly hard on vulnerable people—e.g. the

socially anxious. See for example: “How long does it last? The persistence of the effects of ostracism in the socially anxious”.

So please, be careful in treating niceness and kindness as completely different, especially if your purpose in doing so is to diminish the importance of niceness. (I always feel bad about moralizing so I want to be clear that I am neither especially nice nor kind)

(1)– Another notable feature of this is that our niceness is tested far more frequently than our kindness, so it’s easier to deceive ourselves about our kindness.

What form of ethics is the least subject to rationalising what you wanted to do anyway?

“And now it frightens me, the dreams that I possess

To think I was acting like a believer, when I was just angry and depressed”

-Sufjan Stevens, Ascension

By rationalising ethics I mean finding an “ethical” argument for what you wanted to do anyway. You might think that a disadvantage of consequentialism as compared to deontology is that it’s fairly easy to rationalise consequentialism, to find some argument about extended consequences that suggests that what you wanted to do anyway is actually the right thing to do. But arguably there are plenty of ways to rationalise deontology as well- a complex scheme of duties and exceptions, complex rules around intentions like the doctrine of double effect- all of these provide many degrees of freedom and opportunities for self-interested interpretation.

Note that the space for rationalisation is different in each. In consequentialism it is the complex, fraught and subjective process of estimating the consequences of an action that gives room for rationalisation. In deontology it is the complexity of the system of rules and exceptions in itself. In other words, rationalisation threatens the integrity of consequentialism through questions of fact, and deontology through questions of ethics. It might be possible through careful experimental design to disentangle which form of ethics is, as a matter of fact, more prone to being rationalised(1). This is fascinating because it might give us a hint about whether one form of ethics or the other is what I call practically dominant.

Ethical code E1 is practically dominant over ethical code E2 if and only if, on average, someone trying to live by ethical code E1 will produce more ethical behaviour than if they tried to live by ethical code E2 as judged by the lights of both E1 and E2. If either consequentialism or deontology were grossly more likely to be rationalised, it is possible that the other option might thereby be practically dominant.

Of course one of the many difficulties here is that there is no such thing as consequentialism or deontology generally. Both come in many forms. Still, it's interesting to think about.

Paradox of the book and the robot

This is an old puzzle- I recall reading it in Ted Chiang and I think other places as well. It's a great example of how seemingly reasonable intuitions can lead us astray.

Premise 1. There could exist a book that contains infallibly accurate information about the future.

Premise 2. A robot could read this book.

Premise 3. The book might predict that, at some particular moment, the robot will perform some mundane action, like raising its grasper.

Premise 4. The robot might be programmed to be a perverse robot, in the sense that if anything or anyone makes a prediction about what it will do, it will do the opposite. It might well also be successful in this perversity.

But it seems that premise 3 & 4 can't both be true, if the book is infallible and the robot has read it. It seems that, quite generally if the robot reads the book, the book cannot contain any predictions about what the robot will do voluntarily (assuming the robot's programming remains intact and there are no errors).

But it feels weird doesn't it? Infallible future telling may not exist in our world, but it seems logically possible. If infallible future telling exists, there seems no reason why any agent shouldn't be able to access the results of that future telling, or suffer malfunction or abrogation of its programming if it it does.

The logical paradox here is, at heart, related to, or even identical to, the grand-father killing paradox in time-travel. As there, we must say that in a world where time travel or future telling is possible, any attempt to rewrite events (whether in a fixed past or known future) will always be thwarted.

It's probably also a variant on the "unstoppable force, unmoveable object" paradox.

Two reasons why it's easier to find an audience for writing on the basis of different experiences than novel ideas

Suppose the two basic pitches for a non-fiction bit of media are “I have a different experience” and “I have a novel idea”. Perhaps it’s just my imagination but it seems to me that after about 2008 the balance started tipping from ideas to experiences and the personal essay rose to prominence. Even when ideas are nominally the focus, the identity of the person having them has become more important. This is most obviously true through articles like “here’s what I think of X as a person of Y identity”. However it is also true in subtle ways. For example, the Conversation, and its op-eds written by academics- must be written by academics, and not even people with equivalent experience. A lot of the remaining idea based stuff is done by opinion columnists in legacy publications. Of course there’s Twitter, but that’s very short form.

This is a source of great interest and dismay for me because I’m definitely more of an ideas guy. Like most people I have all sorts of unique, fascinating and grotesque experiences that could be turned into articles, but I don’t like to write about them unless I can find a philosophical angle on them because, for me, it feels a little narcissistic. However, over the years and to my own dismay, I have been drifting further into personal experiences as a source of writing, probably as a result of Darwinian pressure on my behavior(1).

I think there are two reasons why the internet doesn’t have a lot of lot of time for long form latter day op-eds and argumentative essays.

You can immediately tell when someone is bringing new experiences to the table in their writing, or at least experiences different from your own. Working out whether they’re presenting interesting new ideas takes more time. Thus novel ideas lose out in the fierce competition for attention against novel experiences.

There's an ego barrier. We all fancy ourselves intelligent people. We can accept that someone might have different experiences to us, but accepting that they might be a source of insights unknown to us is more galling.

You might be thinking "but hasn't factor two always been true- how does that explain the change"? Answer: Factor two also ties back into factor one. If we had more time to contemplate what someone had to say we might come to accept them as a source of interesting insights, thus overcoming the ego barrier and putting them in the "worth listening to" category- but in this fleeting attention economy we don't.

(1)- There's an essay to be written about how the quantitative feedback (comments, likes, etc.) of this period means that our social behavior is under stronger forces of something like natural selection that ever before. We are quantitatively shaped by clear environmental signals, and behavior, as Skinner noted, can be analogized to an evolutionary process with variation and selection.

The paradox of high expectations: The more you demand, the less you get

In various fairly common situations, demanding more can result in receiving less.

I.

The kinds of situation I am talking about are ubiquitous, but we'll start with employee hiring. Suppose you are running a job search, and are primarily interested in some desirable talent T. Perhaps T is years of experience using some application or programming language.

Now suppose that people come in two varieties, liars and truth tellers. Truth tellers will truthfully (or mostly truthfully) state their level of T. Liars will claim to have whatever level of T they think will make you most likely to hire them. Suppose further that every truth-teller who meets your expectations will apply, and every single liar, regardless of whether they meet your expectations, will apply. Assume the average T value is the same for liars and truth-tellers.

Let's say the population=1100 and 100 of them are liars. Now suppose you set the required level of T at 2 standard deviations above the mean. Assuming a normal distribution, 25 truth tellers will apply, and 100 liars. If you have no way of telling liars apart from truth-tellers, you have an 80% chance of hiring a liar, if you pick from among the applicants who meet the threshold at random.

Meanwhile if you'd set the level of T you were demanding at 1 standard deviation above the mean, you'd have a less than 40% chance of hiring a liar assuming you pick from among the applicants who meet the threshold at random.

If, for example, the value of a candidate to you is equal to their T-value expressed in standard deviations from the mean (positive or negative) you will actually get a lower T

score on average by setting the cutoff at $T=2$, than if you'd just set it at 1.

A lot of readers are probably thinking that our simplifying assumption that employers are no better than chance at spotting liars is too harsh. I have a few things to say to that, viz:

- A) The processes most employers use to determine quality past the resume stage—referee checks and interviews—are in many ways easier to navigate if you're a little bit loose with the truth. It would actually not surprise me if the real process filters out more truth-tellers than liars.
- B) Available empirical evidence suggests that people are generally totally delusional about their ability to spot liars, and spotting liars is actually incredibly difficult.
- C) Even if employers have some skill in filtering liars, if that skill is less than complete, it remains true that, past a certain point, increasing your expectations simply makes it more likely you will get a liar.

One interesting sub-case here is where you are interested in multiple traits, some of which you can check in applicants more easily than others. In these circumstances it may pay to set relatively modest minimum thresholds for the traits you cannot easily check, but adopt a policy of "more=better" for traits that you can accurately check.

II.

Perhaps the most extreme, and comical, variant of this phenomena companies that demand candidates who have more years of experience in programming languages than those languages existed. I don't know if this has ever actually happened, or if it is just a persistent urban legend, but any company that tries this is guaranteed to receive liars. Although job searches are among the most obvious areas in which this paradox arises, it comes up in other areas. For example, seeking quotes, choosing between products and any other domain where the phrase "too good to be true" comes to mind.

When we consider that a pattern of high demands might turn once truth-tellers into

liars, something like this phenomena could explain the flourishing of hypocrisy in some morally rigorous communities. Past a certain point, lying becomes a more workable strategy than actually trying to better yourself.

Demand the very best and you might get much, much less.

Postscript:

One reader suggests that reasoning similar to this is why you should find scientific papers just short of significance considerably more trustworthy.

Another commenter on Reddit suggests that we leave out one very important specification—a third type of person he calls an exaggerator. An exaggerator does not lie to an indefinite degree, but instead exaggerates their qualifications by a fixed amount. If such people are common, it may be prudent to demand a T value of 2, if your real minimum is 1.

This is a really great point, and I guess the overall picture here is that it's difficult to know in the abstract what kinds of demands and criteria are the most likely to work. We've certainly demonstrated that under some conditions, demanding more will lead to less in expectation. In other conditions though this doubtless isn't true. If there is an overall takeaway here, it is that the matter is very complex and unintended consequences abound. I don't really have any algorithmic advice, but it is probably worth thinking through a variety of possibilities and balancing different concerns against each other.

New thought experiments for the backyard metaphysician to try at home

Sam and Finley

Finley loves Sam and Sam loves Finley. Sam hits their head, losing all of their autobiographical memory. However their skills and personality remain as they were. The usual question at this juncture is were you Finley, would you still love Sam? The idea being that if you answer yes, it follows that at least on an emotional level, you believe that Sam is still Sam. Thus, if your feelings are right, continuity of autobiographical memory is not required for continuity of personhood.

I want to ask a different question. It is harder to answer, and perhaps less philosophically illuminating, but still interesting. Suppose that you are Sam instead of Finley. You wake up and the concerned nurse explains to you many things—among them that you have a devoted partner that comes and visits every day. She gives some details of your life together. My two questions are:

1. Do you think that, in this situation, you would immediately, or almost immediately, feel love for Finley—and not just the love you might feel for any kind stranger, but the love of a partner for a partner? If you wouldn't immediately feel love, how quickly do you think it might develop? How likely would it be to develop?
2. Regardless of your answer to the above, do you think you would be obliged to “try to love” Finley. Does the concept of trying to love someone even make sense?

The debate at the end of time

Everyone who has ever died has been raised from the dead in new and immortal bodies. Maybe your resurrectors used some of the technological options I discussed in “Oh Death, Where is the Antidote for thy Sting”, or maybe they used supernatural power, it doesn't matter.

Your resurrectors explain that there is an important quandary—what should be done with the great wrongdoers of history? These have been raised alongside the rest of you. Does Idi Amin deserve an eternity in paradise, should Temujin break bread at the seats of the blessed? They have decided to leave these questions to a democratic decision of every human who has ever lived.

A great debate begins, some arguing they should be absolved with everyone else, some arguing they should be imprisoned for a time, some arguing they should be killed(1), and some arguing for even worse. Who, if anyone, among the resurrected should be punished? How severely should they be punished if at all? How far does this go down the chain of wrongdoing? Should ordinary murderers be punished? Fraudsters? Those guilty of assault? Would you want cultural context to be accepted as a defense? As a mitigating factor? Would the sufferings the wrong doers endured in their own lives count as “time-served”? It falls upon you to take a stance on all these questions, or if not, justify your abstention.

Double trouble

Imagine a world where each body contains two persons each with a very separate personality. Which of these personalities is in the driving seat changes frequently. It's very common for one personality to be cruel while the other is kind, or honest where the other is deceitful.

How would you deal with punishment and criminal justice in this world, given that

punishing a guilty person inevitably also punishes an innocent? What aspects would ethics require us to change? Assume you have a similar level of resourcing to a very well resourced penal system today. Would you try to make prison abolition work? Would you reluctantly accept prisons, but try to greatly minimize their use?

On critical social-technological points

I. Critical Social-Technological Points

I want to introduce the idea of a critical social-technological point (CSTP). A CSTP is a technological discovery or implementation after which the existing hierarchies and ruling class of a society is locked in, in the sense that removing them from power, or even resisting them in any measurably important way, becomes much more difficult. Saying that a technology is a CSTP is different from saying it is an inherently authoritarian technology—A CSTP generally only threatens authoritarianism if it is achieved in an already authoritarian society.

Four examples of possible CSTP's include:

1. Genetic engineering. Genetic engineering could be used to lock in the existence of an unequal power structure by genetically writing in obedience to the underclass, (or it could be used to lock in vibrant democracy by amplifying the human tendency to treat power critically).
2. Surveillance technology. While the mere existence of widespread surveillance has not locked in authoritarianism a la 1984, the creation of artificial intelligences capable of monitoring all surveillance channels for disobedience simultaneously might. If such technologies are successfully deployed by existing illiberal intelligence apparatuses, the capacity for resisting power is greatly reduced, and those apparatuses might become impossible to dislodge.
3. The creation of artificial super intelligence. Writers like Nick Bostrom have worried about an artificial super-intelligence totally out of control of humanity- gaining a decisive strategic advantage. This is indeed terrifying, but only slightly less terrifying is

the prospect of an artificial superintelligence under the control of the wrong segment of humanity gaining a decisive strategic advantage on behalf of its masters.

4. I have previously suggested that, even in the absence of technological unemployment, automation might represent a critical change in the balance of class power and thus a CSTP, permanently removing much of the power of the working class to resist. The idea is that fewer and fewer workers might be necessary for truly essential infrastructure.

II. Past CSTP-lite phenomena

If the thrust of the above is correct, a lot of potential CSTP's are coming up soon. You might rightly wonder if there has been anything like a CSTP in the past? Obviously not in a permanent form because there is, at present, no almost omnipotent ruling class, and probably not as sudden and dramatic as some of the possibilities I outlined above—but something along those lines?

While I am not an antiquarian a few possible candidates include:

1. The invention of agriculture generally.
2. The invention of the ox drawn plow
3. The invention of the chariot
4. The discovery of bronze working

III. What must we do?


The nature of CTSP's is that they block our sight as to what the future looks like past them. Technological advances are inherently hard to predict, as are social confrontations—and the combination forms a cataract in our already cloudy view of the

future. Nonetheless, I think that, cumulatively, between the four possible cases described, we have enough evidence to believe that one or more CTSP's are approaching.

The more power ordinary people have, and the more accountability to which elites are subject, the better our chances are. The creation of a culture of real, conscious democracy in which we jointly and responsibly decide the future—in which bodies like the military and intelligence cannot simply always plead secrecy—in which the spoils of advancing automation is something we all share in, rather than being immiserated by—may be something that must happen soon, if it is to happen at all.

Final bonus

If you're desperate for more philosophy bear content and a massive nerd, check out the draft of my PhD thesis:

 Thesis: Subjective wellbeing as a solution to several key problems of applied welfa...