NVS SOLIDVS

BASIVM VACVVM
XL

PENTAGONA VACVA

Argument and Analysis

1. Eight Short Studies On Excuses
2. Schelling fences on slippery slopes
3. Intellectual Hipsters and Meta-Contrarianism
4. Cardiologists and Chinese Robbers
5. All Debates Are Bravery Debates
6. The Virtue of Silence
7. Proving Too Much
8. Beware Isolated Demands For Rigor
9. Transhumanist Fables
10. ...And I Show You How Deep The Rabbit Hole Goes

Eight Short Studies On Excuses

The Clumsy Game-Player

You and a partner are playing an Iterated Prisoner's Dilemma. Both of you have publicly pre-committed to the tit-for-tat strategy. By iteration 5, you're going happily along, raking up the bonuses of cooperation, when your partner unexpectedly presses the "defect" button.

"Uh, sorry," says your partner. "My finger slipped."

"I still have to punish you just in case," you say. "I'm going to defect next turn, and we'll see how you like it."

"Well," said your partner, "knowing that, I guess I'll defect next turn too, and we'll both lose out. But hey, it was just a slipped finger. By not trusting me, you're costing us both the benefits of one turn of cooperation."

"True", you respond "but if I don't do it, you'll feel free to defect whenever you feel like it, using the 'finger slipped' excuse."

"How about this?" proposes your partner. "I promise to take extra care that my finger won't slip again. You promise that if my finger does slip again, you will punish me terribly, defecting for a bunch of turns. That way, we trust each other again, and we can still get the benefits of cooperation next turn."

You don't believe that your partner's finger really slipped, not for an instant. But the plan still seems like a good one. You accept the deal, and you continue cooperating until the experimenter ends the game.

After the game, you wonder what went wrong, and whether you could have played better. You decide that there was no better way to deal with your partner's "finger-slip" - after all, the plan you enacted gave you maximum possible utility under the circumstances. But you wish that you'd pre-committed, at the beginning, to saying "and I will punish finger slips equally to deliberate defections, so make sure you're careful."

The Lazy Student

You are a perfectly utilitarian school teacher, who attaches exactly the same weight to others' welfare as to your own. You have to have the reports of all fifty students in your class ready by the time midterm grades go out on January 1st. You don't want to have to work during Christmas vacation, so you set a deadline that all reports must be in by December 15th or you won't grade them and the students will fail the class. Oh, and your class is Economics 101, and as part of a class project all your students have to behave as selfish utility-maximizing agents for the year.

It costs your students 0 utility to turn in the report on time, but they gain +1 utility by turning it in late (they enjoy procrastinating). It costs you 0 utility to grade a report turned in before December 15th, but -30 utility to grade one after December 15th.

And students get 0 utility from having their reports graded on time, but get -100 utility from having a report marked incomplete and failing the class.

If you say "There's no penalty for turning in your report after deadline," then the students will procrastinate and turn in their reports late, for a total of +50 utility (1 per student times fifty students). You will have to grade all fifty reports during Christmas break, for a total of - 1500 utility (-30 per report times fifty reports). Total utility is -1450.

So instead you say "If you don't turn in your report on time, I won't grade it." All students calculate the cost of being late, which is +1 utility from procrastinating and -100 from failing the class, and turn in their reports on time. You get all reports graded before Christmas, no students fail the class, and total utility loss is zero. Yay!

Or else - one student comes to you the day after deadline and says "Sorry, I was really tired yesterday, so I really didn't want to come all the way here to hand in my report. I expect you'll grade my report anyway, because I know you to be a perfect utilitarian, and you'd rather take the -30 utility hit to yourself than take the -100 utility hit to me."

You respond "Sorry, but if I let you get away with this, all the other students will turn in their reports late in the summer." She says "Tell you what - our school has procedures for changing a student's previously given grade. If I ever do this again, or if I ever tell anyone else about this, you can change my grade to a fail. Now you know that passing me this one time won't affect anything in the future. It certainly can't affect the past. So you have no reason not to do it." You believe her when she says she'll never tell, but you say "You made this argument because you believed me to be the sort of person who would accept it. In order to prevent other people from making the same argument, I have to be the sort of person who wouldn't accept it. To that end, I'm going to not accept your argument."

The Grieving Student

A second student comes to you and says "Sorry I didn't turn in my report yesterday. My mother died the other day, and I wanted to go to her funeral."

You say "Like all economics professors, I have no soul, and so am unable to sympathize with your loss. Unless you can make an argument that would apply to all rational actors in my position, I can't grant you an extension."

She says "If you did grant this extension, it wouldn't encourage other students to turn in their reports late. The other students would just say 'She got an extension because her mother died'. They know they won't get extensions unless they kill their own mothers, and even economics students aren't that evil. Further, if you don't grant the extension, it won't help you get more reports in on time. Any student would rather attend her mother's funeral than pass a course, so you won't be successfully motivating anyone else to turn in their reports early."

You think for a while, decide she's right, and grant her an extension on her report.

The Sports Fan

A third student comes to you and says "Sorry I didn't turn in my report yesterday. The Bears' big game was on, and as I've told you before, I'm a huge Bears fan. But don't worry! It's very rare that there's a game on this important, and not many students here are sports fans anyway. You'll probably never see a student with this exact

excuse again. So in a way, it's not that different from the student here just before me, the one whose mother died."

You respond "It may be true that very few people will be able to say both that they're huge Bears fans, and that there's a big Bears game on the day before the report comes due. But by accepting your excuse, I establish a precedent of accepting excuses that are *approximately this good*. And there are many other excuses approximately as good as yours. Maybe someone's a big soap opera fan, and the season finale is on the night before the deadline. Maybe someone loves rock music, and there's a big rock concert on. Maybe someone's brother is in town that week. Practically anyone can come up with an excuse as good as yours, so if I accept your late report, I have to accept everyone's.

"The student who was here before you, that's different. We, as a society, already have an ordering in which a family member's funeral is one of the most important things around. By accepting her excuse, I'm establishing a precedent of accepting any excuse approximately that good, but almost no one will ever have an excuse that good. Maybe a few people who are really sick, someone struggling with a divorce or a breakup, that kind of thing. Not the hordes of people who will be coming to me if I give you your exemption.

The Murderous Husband

You are the husband of a wonderful and beautiful lady whom you love very much - and whom you just found in bed with another man. In a rage, you take your hardcover copy of Introduction To Game Theory and knock him over the head with it, killing him instantly (it's a pretty big book).

At the murder trial, you plead to the judge to let you go free. "Society needs to lock up murderers, as a general rule. After all, they are dangerous people who cannot be allowed to walk free. However, I only killed that man because he was having an affair with my wife. In my place, anyone would have done the same. So the crime has no bearing on how likely I am to murder someone else. I'm not a risk to anyone who isn't having an affair with my wife, and after this incident I plan to divorce and live the rest of my days a bachelor. Therefore, you have no need to deter me from future murders, and can safely let me go free."

The judge responds: "You make a convincing argument, and I believe that you will never kill anyone else in the future. However, other people will one day be in the position you were in, where they walk in on their wives having an affair. Society needs to have a credible pre-commitment to punishing them if they succumb to their rage, in order to deter them from murder."

"No," you say, "I understand your reasoning, but it won't work. If you've never walked in on your wife having an affair, you can't possibly understand the rage. No matter how bad the deterrent was, you'd still kill the guy."

"Hm," says the judge. "I'm afraid I just can't believe anyone could ever be quite that irrational. But I see where you're coming from. I'll give you a lighter sentence."

The Bellicose Dictator

You are the dictator of East Examplestan, a banana republic subsisting off its main import, high quality hypothetical scenarios. You've always had it in for your ancestral enemy, West Examplestan, but the UN has made it clear that any country in your region that aggressively invades a neighbor will be severely punished with sanctions and possible enforced "regime change." So you decide to leave the West alone for the time being.

One day, a few West Examplestanis unintentionally wander over your unmarked border while prospecting for new scenario mines. You immediately declare it a "hostile incursion" by "West Examplestani spies", declare war, and take the Western capital in a sneak attack.

The next day, Ban Ki-moon is on the phone, and he sounds angry. "I thought we at the UN had made it perfectly clear that countries can't just invade each other anymore!"

"But didn't you read our propaganda mouthpi...ahem, official newspaper? We didn't *just* invade. We were responding to Western aggression!"

"Balderdash!" says the Secretary-General. "Those were a couple of lost prospectors, and you know it!"

"Well," you say. "Let's consider your options. The UN needs to make a credible pre-commitment to punish aggressive countries, or everyone will invade their weaker neighbors. And you've got to follow through on your threats, or else the pre-commitment won't be credible anymore. But you don't actually like following through on your threats. Invading rogue states will kill a lot of people on both sides and be politically unpopular, and sanctions will hurt your economy *and* lead to heart-rending images of children starving. What you'd really like to do is let us off, but in a way that doesn't make other countries think they'll get off too.

"Luckily, we can make a credible story that we were following international law. Sure, it may have been stupid of us to mistake a few prospectors for an invasion, but there's no international law against being stupid. If you dismiss us as simply misled, you don't have to go through the trouble of punishing us, and other countries won't think they can get away with anything.

"Nor do you need to live in fear of us doing something like this again. We've already demonstrated that we won't go to war without a casus belli. If other countries can refrain from giving us one, they have nothing to fear."

Ban Ki-moon doesn't believe your story, but the countries that would bear the economic brunt of the sanctions and regime change decide they believe it just enough to stay uninvolved.

The Peyote-Popping Native

You are the governor of a state with a large Native American population. You have banned all mind-altering drugs, with the honorable exceptions of alcohol, tobacco, caffeine, and several others, because you are a red-blooded American who believes that they would drive teenagers to commit crimes.

A representative of the state Native population comes to you and says: "Our people have used peyote religiously for hundreds of years. During this time, we haven't become addicted or committed any crimes. Please grant us a religious exemption under the First Amendment to continue practicing our ancient rituals." You agree.

A leader of your state's atheist community breaks into your office via the ventilation systems (because seriously, how else is an atheist leader going to get access to a state governor?) and says: "As an atheist, I am offended that you grant exemptions to your anti-peyote law for religious reasons, but not for, say, recreational reasons. This is unfair discrimination in favor of religion. The same is true of laws that say Sikhs can wear turbans in school to show support for God, but my son can't wear a baseball cap in school to show support for the Yankees. Or laws that say Muslims can get time off state jobs to pray five times a day, but I can't get time off my state job for a cigarette break. Or laws that say state functions will include special kosher meals for Jews, but not special pasta meals for people who really like pasta."

You respond "Although my policies may seem to be saying religion is more important than other potential reasons for breaking a rule, one can make a non-religious case justifying them. One important feature of major world religions is that their rituals have been fixed for hundreds of years. Allowing people to break laws for religious reasons makes religious people very happy, but does not weaken the laws. After all, we all know the few areas in which the laws of the major US religions as they are currently practiced conflict with secular law, and none of them are big deals. So the general principle 'I will allow people to break laws if it is necessary to established and well-known religious rituals" is relatively low-risk and makes people happy without threatening the concept of law in general. But the general principle 'I will allow people to break laws for recreational reasons' is *very* high risk, because it's sufficient justification for almost anyone breaking any law."

"I would love to be able to serve everyone the exact meal they most wanted at state dinners. But if I took your request for pasta because you liked pasta, I would have to follow the general principle of giving everyone the meal they most like, which would be prohibitively expensive. By giving Jews kosher meals, I can satisfy a certain particularly strong preference without being forced to satisfy anyone else's."

The Well-Disguised Atheist

The next day, the atheist leader comes in again. This time, he is wearing a false mustache and sombrero. "I represent the Church of Driving 50 In A 30 Mile Per Hour Zone," he says. "For our members, going at least twenty miles per hour over the speed limit is considered a sacrament. Please grant us a religious exemption to traffic laws."

You decide to play along. "How long has your religion existed, and how many people do you have?" you ask.

"Not very long, and not very many people," he responds.

"I see," you say. "In that case, you're a cult, and not a religion at all. Sorry, we don't deal with cults."

"What, exactly, is the difference between a cult and a religion?"

"The difference is that cults have been formed recently enough, and are small enough, that we are suspicious of them existing for the purpose of taking advantage of the special place we give religion. Granting an exemption for your cult would challenge the credibility of our pre-commitment to punish people who break the law, because it would mean anyone who wants to break a law could just found a cult dedicated to it."

"How can my cult become a real religion that deserves legal benefits?"

"You'd have to become old enough and respectable enough that it becomes implausible that it was created for the purpose of taking advantage of the law."

"That sounds like a lot of work."

"Alternatively, you could try writing awful science fiction novels and hiring a ton of lawyers. I hear that also works these days."

Conclusion

In all these stories, the first party wants to credibly pre-commit to a rule, but also has incentives to forgive other people's deviations from the rule. The second party breaks the rules, but comes up with an excuse for why its infraction should be forgiven.

The first party's response is based not only on whether the person's excuse is believable, not even on whether the person's excuse is morally valid, but on whether the excuse can be accepted without straining the credibility of their previous pre-commitment.

The general principle is that by accepting an excuse, a rule-maker is also committing themselves to accepting all equally good excuses in the future. There are some exceptions - accepting an excuse in private but making sure no one else ever knows, accepting an excuse once with the express condition that you will never accept any other excuses - but to some degree these are devil's bargains, as anyone who can predict you will do this can take advantage of you.

These stories give an idea of excuses different from the one our society likes to think it uses, namely that it accepts only excuses that are true and that reflect well upon the character of the person giving the excuse. I'm not saying that the common idea of excuses doesn't have value - but I think the game theory view also has some truth to it. I also think the game theoretic view can be useful in cases where the common view fails. It can inform cases in law, international diplomacy, and politics where a tool somewhat stronger than the easily-muddled common view is helpful.

Schelling fences on slippery slopes

Slippery slopes are themselves a slippery concept. Imagine trying to explain them to an alien:

"Well, we right-thinking people are quite sure that the Holocaust happened, so banning Holocaust denial would shut up some crackpots and improve the discourse. But it's one step on the road to things like banning unpopular political positions or religions, and we right-thinking people oppose that, so we won't ban Holocaust denial."

And the alien might well respond: "But you could just ban Holocaust denial, but not ban unpopular political positions or religions. Then you right-thinking people get the thing you want, but not the thing you don't want."

This post is about some of the replies you might give the alien.

Abandoning the Power of Choice

This is the boring one without any philosophical insight that gets mentioned only for completeness' sake. In this reply, giving up a certain point risks losing the ability to decide whether or not to give up other points.

For example, if people gave up the right to privacy and allowed the government to monitor all phone calls, online communications, and public places, then if someone launched a military coup, it would be very difficult to resist them because there would be no way to secretly organize a rebellion. This is also brought up in arguments about gun control a lot.

I'm not sure this is properly thought of as a slippery slope argument at all. It seems to be a more straightforward "Don't give up useful tools for fighting tyranny" argument.

The Legend of Murder-Gandhi

Previously on Less Wrong's *The Adventures of Murder-Gandhi*: Gandhi is offered a pill that will turn him into an unstoppable murderer. He refuses to take it, because in his current incarnation as a pacifist, he doesn't want others to die, and he knows that would be a consequence of taking the pill. Even if we offered him $1 million to take the pill, his abhorrence of violence would lead him to refuse.

But suppose we offered Gandhi $1 million to take a different pill: one which would decrease his reluctance to murder by 1%. This sounds like a pretty good deal. Even a person with 1% less reluctance to murder than Gandhi is still pretty pacifist and not likely to go killing anybody. And he could donate the money to his favorite charity and perhaps save some lives. Gandhi accepts the offer.

Now we iterate the process: every time Gandhi takes the 1%-more-likely-to-murder-pill, we offer him another $1 million to take the same pill again.

Maybe original Gandhi, upon sober contemplation, would decide to accept $5 million to become 5% less reluctant to murder. Maybe 95% of his original pacifism is the only level at which he can be *absolutely sure* that he will still pursue his pacifist ideals.

Unfortunately, original Gandhi isn't the one making the choice of whether or not to take the 6th pill. 95%-Gandhi is. And 95% Gandhi doesn't care *quite* as much about pacifism as original Gandhi did. He still doesn't want to become a murderer, but it wouldn't be a disaster if he were just 90% as reluctant as original Gandhi, that stuck-up goody-goody.

What if there were a general principle that each Gandhi was comfortable with Gandhis 5% more murderous than himself, but no more? Original Gandhi would start taking the pills, hoping to get down to 95%, but 95%-Gandhi would start taking five more, hoping to get down to 90%, and so on until he's rampaging through the streets of Delhi, killing everything in sight.

Now we're tempted to say Gandhi shouldn't even take the first pill. But this also seems odd. Are we really saying Gandhi shouldn't take what's basically a free million dollars to turn himself into 99%-Gandhi, who might well be nearly indistinguishable in his actions from the original?

Maybe Gandhi's best option is to "fence off" an area of the slippery slope by establishing a Schelling point - an arbitrary point that takes on special value as a dividing line. If he can hold himself to the precommitment, he can maximize his winnings. For example, original Gandhi could swear a mighty oath to take only five pills - or if he didn't trust even his own legendary virtue, he could give all his most valuable possessions to a friend and tell the friend to destroy them if he took more than five pills. This would commit his future self to stick to the 95% boundary (even though that future self is itching to try to the same precommitment strategy to stick to its own 90% boundary).

Real slippery slopes will resemble this example if, each time we change the rules, we also end up changing our opinion about how the rules should be changed. For example, I think the Catholic Church may be working off a theory of "If we give up this traditional practice, people will lose respect for tradition and want to give up even more traditional practices, and so on."

Slippery Hyperbolic Discounting

One evening, I start playing *Sid Meier's Civilization* (IV, if you're wondering - V is terrible). I have work tomorrow, so I want to stop and go to sleep by midnight.

At midnight, I consider my alternatives. For the moment, I feel an urge to keep playing Civilization. But I know I'll be miserable tomorrow if I haven't gotten enough sleep. Being a hyperbolic discounter, I value the next ten minutes a lot, but after that the curve becomes pretty flat and maybe I don't value 12:20 much more than I value the next morning at work. Ten minutes' sleep here or there doesn't make any difference. So I say: "I will play Civilization for ten minutes - 'just one more turn' - and then I will go to bed."

Time passes. It is now 12:10. Still being a hyperbolic discounter, I value the next ten minutes a lot, and subsequent times much less. And so I say: I will play until 12:20, ten minutes sleep here or there not making much difference, and then sleep.

And so on until my empire bestrides the globe and the rising sun peeps through my windows.

This is pretty much the same process described above with Murder-Gandhi except that here the role of the value-changing pill is played by time and my own tendency to discount hyperbolically.

The solution is the same. If I consider the problem early in the evening, I can precommit to midnight as a nice round number that makes a good Schelling point. Then, when deciding whether or not to play after midnight, I can treat my decision not as "Midnight or 12:10" - because 12:10 will always win *that* particular race - but as "Midnight or abandoning the only credible Schelling point and probably playing all night", which will be sufficient to scare me into turning off the computer.

(if I consider the problem at 12:01, I may be able to precommit to 12:10 if I am especially good at precommitments, but it's not a very natural Schelling point and it might be easier to say something like "as soon as I finish this turn" or "as soon as I discover this technology").

Coalitions of Resistance

Suppose you are a Zoroastrian, along with 1% of the population. In fact, along with Zoroastrianism your country has fifty other small religions, each with 1% of the population. 49% of your countrymen are atheist, and hate religion with a passion.

You hear that the government is considering banning the Taoists, who comprise 1% of the population. You've never liked the Taoists, vile doubters of the light of Ahura Mazda that they are, so you go along with this. When you hear the government wants to ban the Sikhs and Jains, you take the same tack.

But now you are in the unfortunate situation described by Martin Niemoller:

> *First they came for the socialists, and I did not speak out, because I was not a socialist.*
> *Then they came for the trade unionists, and I did not speak out, because I was not a trade unionist.*
> *Then they came for the Jews, and I did not speak out, because I was not a Jew.*
> *Then they came for me, but we had already abandoned the only defensible Schelling point*

With the banned Taoists, Sikhs, and Jains no longer invested in the outcome, the 49% atheist population has enough clout to ban Zoroastrianism and anyone else they want to ban. The better strategy would have been to have all fifty-one small religions form a coalition to defend one another's right to exist. In this toy model, they could have done so in an ecumenial congress, or some other literal strategy meeting.

But in the real world, there aren't fifty-one well-delineated religions. There are billions of people, each with their own set of opinions to defend. It would be impractical for everyone to physically coordinate, so they have to rely on Schelling points.

In the original example with the alien, I cheated by using the phrase "right-thinking people". In reality, figuring out who qualifies to join the Right-Thinking People Club is half the battle, and everyone's likely to have a different opinion on it. So far, the practical solution to the coordination problem, the "only defensible Schelling point", has been to just have everyone agree to defend everyone else without worrying whether they're right-thinking or not, and this is easier than trying to coordinate room for exceptions like Holocaust deniers. Give up on the Holocaust deniers, and no one

else can be sure what other Schelling point you've committed to, if any...

...unless they can. In parts of Europe, they've banned Holocaust denial for years and everyone's been totally okay with it. There are also a host of other well-respected exceptions to free speech, like shouting "fire" in a crowded theater. Presumably, these exemptions are protected by tradition, so that they have become new Schelling points there, or are else so obvious that everyone except Holocaust deniers is willing to allow a special Holocaust denial exception without worrying it will impact their own case.

Summary

Slippery slopes legitimately exist wherever a policy not only affects the world directly, but affects people's willingness or ability to oppose future policies. Slippery slopes can sometimes be avoided by establishing a "Schelling fence" - a Schelling point that the various interest groups involved - or yourself across different values and times - make a credible precommitment to defend.

Intellectual Hipsters and Meta-Contrarianism

Related to: Why Real Men Wear Pink, That Other Kind of Status, Pretending to be Wise, The "Outside The Box" Box

WARNING: Beware of things that are fun to argue -- Eliezer Yudkowsky

Science has inexplicably failed to come up with a precise definition of "hipster", but from my limited understanding a hipster is a person who deliberately uses unpopular, obsolete, or obscure styles and preferences in an attempt to be "cooler" than the mainstream. But why would being deliberately uncool be cooler than being cool?

As previously discussed, in certain situations refusing to signal can be a sign of high status. Thorstein Veblen invented the term "conspicuous consumption" to refer to the showy spending habits of the nouveau riche, who unlike the established money of his day took great pains to signal their wealth by buying fast cars, expensive clothes, and shiny jewelery. Why was such flashiness common among new money but not old? Because the old money was so secure in their position that it never even occurred to them that they might be confused with poor people, whereas new money, with their lack of aristocratic breeding, worried they might be mistaken for poor people if they didn't make it blatantly obvious that they had expensive things.

The old money might have started off not buying flashy things for pragmatic reasons - they didn't need to, so why waste the money? But if F. Scott Fitzgerald is to be believed, the old money actively cultivated an air of superiority to the nouveau riche and their conspicuous consumption; not buying flashy objects becomes a matter of principle. This makes sense: the nouveau riche need to differentiate themselves from the poor, but the old money need to differentiate themselves from the nouveau riche.

This process is called countersignaling, and one can find its telltale patterns in many walks of life. Those who study human romantic attraction warn men not to "come on too strong", and this has similarities to the nouveau riche example. A total loser might come up to a woman without a hint of romance, promise her nothing, and demand sex. A more sophisticated man might buy roses for a woman, write her love poetry, hover on her every wish, et cetera; this signifies that he is not a total loser. But the most desirable men may deliberately avoid doing nice things for women in an attempt to signal they are so high status that they don't need to. The average man tries to differentiate himself from the total loser by being nice; the extremely attractive man tries to differentiate himself from the average man by not being especially nice.

In all three examples, people at the top of the pyramid end up displaying characteristics similar to those at the bottom. Hipsters deliberately wear the same clothes uncool people wear. Families with old money don't wear much more jewelry than the middle class. And very attractive men approach women with the same lack of subtlety a total loser would use.[1]

If politics, philosophy, and religion are really about signaling, we should expect to find countersignaling there as well.

Pretending To Be Wise

Let's go back to Less Wrong's long-running discussion on death. Ask any five year old child, and ey can tell you that death is bad. Death is bad because it kills you. There is nothing subtle about it, and there does not need to be. Death universally seems bad to pretty much everyone on first analysis, and what it seems, it is.

But as has been pointed out, along with the gigantic cost, death does have a few small benefits. It lowers overpopulation, it allows the new generation to develop free from interference by their elders, it provides motivation to get things done quickly. Precisely because these benefits are so much smaller than the cost, they are hard to notice. It takes a particularly subtle and clever mind to think them up. Any idiot can tell you why death is bad, but it takes a very particular sort of idiot to believe that death might be good.

So pointing out this contrarian position, that death has some benefits, is potentially a signal of high intelligence. It is not a very reliable signal, because once the first person brings it up everyone can just copy it, but it is a cheap signal. And to the sort of person who might not be clever enough to come up with the benefits of death themselves, and only notices that wise people seem to mention death can have benefits, it might seem super extra wise to say death has lots and lots of great benefits, and is really quite a good thing, and if other people should protest that death is bad, well, that's an opinion a five year old child could come up with, and so clearly that person is no smarter than a five year old child. Thus Eliezer's title for this mentality, "Pretending To Be Wise".

If dwelling on the benefits of a great evil is not your thing, you can also pretend to be wise by dwelling on the costs of a great good. All things considered, modern industrial civilization - with its advanced technology, its high standard of living, and its lack of typhoid fever - is pretty neat. But modern industrial civilization also has many costs: alienation from nature, strains on the traditional family, the anonymity of big city life, pollution and overcrowding. These are real costs, and they are certainly worth taking seriously; nevertheless, the crowds of emigrants trying to get from the Third World to the First, and the lack of any crowd in the opposite direction, suggest the benefits outweigh the costs. But in my estimation - and speak up if you disagree - people spend a lot more time dwelling on the negatives than on the positives, and most people I meet coming back from a Third World country have to talk about how much more authentic their way of life is and how much we could learn from them. This sort of talk sounds Wise, whereas talk about how nice it is to have buses that don't break down every half mile sounds trivial and selfish..

So my hypothesis is that if a certain side of an issue has very obvious points in support of it, and the other side of an issue relies on much more subtle points that the average person might not be expected to grasp, then adopting the second side of the issue will become a signal for intelligence, even if that side of the argument is wrong.

This only works in issues which are so muddled to begin with that there is no fact of the matter, or where the fact of the matter is difficult to tease out: so no one tries to signal intelligence by saying that 1+1 equals 3 (although it would not surprise me to find a philosopher who says truth is relative and this equation is a legitimate form of discourse).

Meta-Contrarians Are Intellectual Hipsters

A person who is somewhat upper-class will conspicuously signal eir wealth by buying difficult-to-obtain goods. A person who is very upper-class will conspicuously signal that ey feels no need to conspicuously signal eir wealth, by deliberately not buying difficult-to-obtain goods.

A person who is somewhat intelligent will conspicuously signal eir intelligence by holding difficult-to-understand opinions. A person who is very intelligent will conspicuously signal that ey feels no need to conspicuously signal eir intelligence, by deliberately not holding difficult-to-understand opinions.

According to the survey, the average IQ on this site is around 145[2]. People on this site differ from the mainstream in that they are more willing to say death is bad, more willing to say that science, capitalism, and the like are good, and less willing to say that there's some deep philosophical sense in which 1+1 = 3. That suggests people around that level of intelligence have reached the point where they no longer feel it necessary to differentiate themselves from the sort of people who aren't smart enough to understand that there might be side benefits to death. Instead, they are at the level where they want to differentiate themselves from the somewhat smarter people who think the side benefits to death are great. They are, basically, meta-contrarians, who counter-signal by holding opinions contrary to those of the contrarians' signals. And in the case of death, this cannot but be a good thing.

But just as contrarians risk becoming too contrary, moving from "actually, death has a few side benefits" to "DEATH IS GREAT!", meta-contrarians are at risk of becoming too meta-contrary.

All the possible examples here are controversial, so I will just take the least controversial one I can think of and beg forgiveness. A naive person might think that industrial production is an absolute good thing. Someone smarter than that naive person might realize that global warming is a strong negative to industrial production and desperately needs to be stopped. Someone even smarter than that, to differentiate emself from the second person, might decide global warming wasn't such a big deal after all, or doesn't exist, or isn't man-made.

In this case, the contrarian position happened to be right (well, maybe), and the third person's meta-contrariness took em further from the truth. I do feel like there are more global warming skeptics among what Eliezer called "the atheist/libertarian/technophile/sf-fan/early-adopter/programmer empirical cluster in personspace" than among, say, college professors.

In fact, very often, the uneducated position of the five year old child may be deeply flawed and the contrarian position a necessary correction to those flaws. This makes meta-contrarianism a very dangerous business.

Remember, most everyone hates hipsters.

Without meaning to imply anything about whether or not any of these positions are correct or not[3], the following triads come to mind as connected to an uneducated/contrarian/meta-contrarian divide:

- KKK-style racist / politically correct liberal / "but there are scientifically proven genetic differences"

- misogyny / women's rights movement / men's rights movement
- conservative / liberal / libertarian[4]
- herbal-spiritual-alternative medicine / conventional medicine / Robin Hanson
- don't care about Africa / give aid to Africa / don't give aid to Africa
- Obama is Muslim / Obama is obviously not Muslim, you idiot / Patri Friedman[5]

What is interesting about these triads is not that people hold the positions (which could be expected by chance) but that people get deep personal satisfaction from arguing the positions even when their arguments are unlikely to change policy[6] - and that people identify with these positions to the point where arguments about them can become personal.

If meta-contrarianism is a real tendency in over-intelligent people, it doesn't mean they should immediately abandon their beliefs; that would just be meta-meta-contrarianism. It means that they need to recognize the meta-contrarian tendency within themselves and so be extra suspicious and careful about a desire to believe something contrary to the prevailing contrarian wisdom, especially if they really enjoy doing so.

Footnotes

1) But what's really interesting here is that people at each level of the pyramid don't just follow the customs of their level. They enjoy following the customs, it makes them feel good to talk about how they follow the customs, and they devote quite a bit of energy to insulting the people on the other levels. For example, old money call the nouveau riche "crass", and men who don't need to pursue women call those who do "chumps". Whenever holding a position makes you feel superior and is fun to talk about, that's a good sign that the position is not just practical, but signaling related.

2) There is no need to point out just how unlikely it is that such a number is correct, nor how unscientific the survey was.

3) One more time: *the fact that those beliefs are in an order does not mean some of them are good and others are bad*. For example, "5 year old child / pro-death / transhumanist" is a triad, and "warming denier / warming believer / warming skeptic" is a triad, but I personally support 1+3 in the first triad and 2 in the second. You can't evaluate the truth of a statement by its position in a signaling game; otherwise you could use human psychology to figure out if global warming is real!

4) This is my solution to the eternal question of why libertarians are always more hostile toward liberals, even though they have just about as many points of real disagreement with the conservatives.

5) To be fair to Patri, he admitted that those two posts were "trolling", but I think the fact that he derived so much enjoyment from trolling in that particular way is significant.

6) Worth a footnote: I think in a lot of issues, the original uneducated position has disappeared, or been relegated to a few rednecks in some remote corner of the world, and so meta-contrarians simply look like contrarians. I think it's important to keep the terminology, because most contrarians retain a psychology of feeling like they are being contrarian, even after they are the new norm. But my only evidence for this is introspection, so it might be false.

Cardiologists and Chinese Robbers

I.

It takes a special sort of person to be a cardiologist. This is not always a good thing.

You may have read about one or another of the "cardiologist caught falsifying test results and performing dangerous unnecessary surgeries to make more money" stories, but you might not have realized just how common it really is. Maryland cardiologist performs over 500 dangerous unnecessary surgeries to make money. Unrelated Maryland cardiologist performs another 25 in a separate incident. California cardiologist does "several hundred" dangerous unnecessary surgeries and gets raided by the FBI. Philadelphia cardiologist, same. North Carolina cardiologist, same. 11 Kentucky cardiologists, same. Actually just a couple of miles from my own hospital, a Michigan cardiologist was found to have done $4 million worth of the same. Etc, etc, etc.

My point is not just about the number of cardiologists who perform dangerous unnecessary surgeries for a quick buck. It's not even just about the cardiology insurance fraud, cardiology kickback schemes, or cardiology research data falsification conspiracies. That could all just be attributed to some distorted incentives in cardiology as a field. My point is that it takes a special sort of person to be a cardiologist.

Consider the sexual harassment. Head of Yale cardiology department fired for sexual harassment with "rampant bullying". Stanford cardiologist charged with sexually harassing students. Baltimore cardiologist found guilty of sexual harassment. LA cardiologist fined $200,000 for groping med tech. Three different Pennsylvania cardiologists sexually harassing the same woman. Arizona cardiologist suspended on 19 (!) different counts of sexual abuse. One of the "world's leading cardiologists" fired for sending pictures of his genitals to a female friend. New York cardiologist in trouble for refusing to pay his $135,000 bill at a strip club. Manhattan cardiologist taking naked pictures of patients, then using them to sexually abuse employees. New York cardiologist secretly installs spycam in office bathroom. Just to shake things up, a Florida cardiologist was falsely accused of sexual harassment as part of feud with another cardiologist.

And yeah, you can argue that if you put high-status men in an office with a lot of subordinates, sexual harassment will be depressingly common just as a result of the environment. But there's also the Texas cardiologist who pled guilty to child molestation. The California cardiologist who killed a two-year-old kid. The author of one of the world's top cardiology textbooks arrested on charges Wikipedia describes only as "related to child pornography and cocaine".

Then it gets weird. Did you about the Australian cardiologist who is fighting against extradition to Uganda, where he is accused of "terrorism, aggravated robbery and murdering seven people"? What about the Long Island cardiologist who hired a hitman to kill a rival cardiologist, ahd who was also for some reason looking for "enough explosives to blow up a building"?

Like I said, it takes a special sort of person.

II.

Given the recent discussion of media bias here, I wanted to bring up Alyssa Vance's "Chinese robber fallacy", which she describes as:

> ..where you use a generic problem to attack a specific person or group, even though other groups have the problem just as much (or even more so).

For example, if you don't like Chinese people, you can find some story of a Chinese person robbing someone, and claim that means there's a big social problem with Chinese people being robbers.

I originally didn't find this too interesting. It sounds like the same idea as plain old stereotyping, something we think about often and are carefully warned to avoid.

But after re-reading the post, I think the argument is more complex. There are over a billion Chinese people. If even one in a thousand is a robber, you can provide one million examples of Chinese robbers to appease the doubters. Most people think of stereotyping as "Here's one example I heard of where the out-group does something bad," and then you correct it with "But we can't generalize about an entire group just from one example!" It's less obvious that you may be able to provide literally one million examples of your false stereotype and still have it be a false stereotype. If you spend twelve hours a day on the task and can describe one crime every ten seconds, you can spend four months doing nothing but providing examples of burglarous Chinese – and still have absolutely no point.

If we're really concerned about media bias, we need to think about Chinese Robber Fallacy as one of the media's strongest weapons. There are lots of people – 300 million in America alone. No matter what point the media wants to make, there will be hundreds of salient examples. No matter how low-probability their outcome of interest is, they will never have to stop covering it if they don't want to.

This has briefly gotten some coverage in the form of "the war on police". As per AEI:

> Is there a "war on police" in America today? Most Americans think so, and that's understandable given all of the media coverage of that topic. A Google news search finds 32,000 results for the phrase "war on cops" and another 12,100 results for "war on police," with sensational headlines like "America's War on Cops Intensifies" and "Bratton Warns of Tough Times Ahead Due to 'War on Cops'." A recent Rasmussen poll found that 58% of likely US voters answered "Yes" to the question "Is there a war on police in America today?" and only 27% disagreed. But data on police shootings in America that were reported last week by The Guardian tell a much different story of increasing police safety.
>
> According to data available from the "Officer Down Memorial Page" on the annual number of non-accidental, firearm-related police fatalities, 2015 is on track to be the safest year for law enforcement in the US since 1887 (except for a slightly safer year in 2013), more than 125 years ago. And adjusted for the country's growing population, the years 2013 and 2015 will be the two safest years for police in US history, measured by the annual number of firearm-related police fatalities per 1 million people.

When politically convenient, it is easy to make Americans believe in a war on police simply by better coverage of existing murders of police officers. Given that America is a big country with very many police, even a low base rate will provide many lurid police-officer-murder stories – by my calculation, two murders a week even if officers are killed only at the same rate as everyone else. While covering these is a legitimate

decision, it can be deceptive unless it's framed in terms of things like whether the rate has gone up or down, whether the rate is higher or lower for the group involved than the base rate in the population, and it still seems scary when you explicitly calculate the rate.

But a Chomskian analysis would ask whether the talk of a "war on cops" is really a uniquely bad example of journalistic malpractice, or whether it is bog-standard journalistic malpractice which is unique only in being called out this time instead of allowed to pass.

Let's stick with coverage of police for consistency's sake. I've made a very similar argument before regarding claims of racist police shootings (see Part D here), but let's avoid that particular rabbit hole and consider a broader and more unsettling point. We all hear anecdotes about terrible police brutality. Suppose, in fact, that we've heard exactly X stories. Given that there are about 100,000 police officers in the US, is X consistent with the problem being systemic and dire, or with the problem being relatively limited?

I mean, it's hard to say. Quick Fermi calculation: if I can think of about one horrible story of police brutality a week, and assume there are fifty that aren't covered for every one that is, then per year that makes…

But wait – what if I told you that number was a lie, and there were actually 500,000 police officers in the US? Suddenly the rate of police brutality has decreased five times from what it was a second ago. If you previously believed that there were 100,000 police officers, and that the police brutality rate was shameful but that decreasing the rate to only one-fifth its previous level would count as a victory, well, now you can declare victory.

What if I told you the 500,000 number is also a lie, and it's actually way more cops than that? Do you have any idea at all how many police there are? Shouldn't you at least have an order-of-magnitude estimate of what the police brutality rate is before deciding if it's too high or not? What if I told you the real number was a million cops? Five million cops? Ten million? That's a hundred times the original estimate of 100,000 – shouldn't learning that the police brutality rate is only 1% of what you originally estimated (or, going the other direction, 10,000% of that) change your opinion in some way?

(No, I won't tell you how many cops there actually are. Look it up.)

I feel this way about a lot of things. The media is always giving us stories of how tech nerds are sexist in some way or another. But we may suspect they want to push that line regardless of whether it's true. How many tech nerds are there? A million? Ten million? How many lurid stories about harassment in Silicon Valley have you heard? Do we know if this is higher or lower than the base rate for similar industries? Whether it's going up or down? What it would look like if we actually had access to the per person rates?

By now you've probably figured out the gimmick, but just to come totally clean – cardiologists are wonderful people who as far as I know are no less ethical than any other profession. I chose to pick on them at random – well, not quite random, one of them yelled at me the other day because apparently contacting the cardiologist on call late at night just because your patient is having a serious heart-related emergency is some kind of huge medical faux pas. I don't think anyone has ever

claimed that there’s any general issue with cardiologists, and as far as I know there’s no evidence for such.

If you read Part I of this post and found yourself nodding along, thinking “Wow, cardiologists are real creeps, there must be serious structural problems in the cardiology profession, something must be done about them,” consider it evidence that a sufficiently motivated individual – especially a journalist! – can make you feel that way about any group.

All Debates Are Bravery Debates

"I don't practice what I preach because I'm not the kind of person I'm preaching to."

— Bob Dobbs

I.

I read Atlas Shrugged probably about a decade ago, and felt turned off by its promotion of selfishness as a moral ideal. I thought that was basically just being a jerk. After all, if there's one thing the world doesn't need (I thought) it's more selfishness.

Then I talked to a friend who told me Atlas Shrugged had changed his life. That he'd been raised in a really strict family that had told him that ever enjoying himself was selfish and made him a bad person, that he had to be working at every moment to make his family and other people happy or else let them shame him to pieces. And the revelation that it was sometimes okay to consider your own happiness gave him the strength to stand up to them and turn his life around, while still keeping the basic human instinct of helping others when he wanted to and he felt they deserved it (as, indeed, do Rand characters).

II.

The religious and the irreligious alike enjoy making fun of Reddit's r/atheism, which combines an extreme strawmanning of religious positions with childish insults and distasteful triumphalism. Recently the moderators themselves have become a bit embarrassed by it and instituted some rules intended to tone things down, leading to some of the most impressive Internet drama I have ever seen. In its midst, some people started talking about what the old strawmanning triumphalist r/atheism meant to them (see for example here).

A lot of them were raised in religious families where they would have been disowned if they had admitted to their atheism. Some of them were disowned for admitting to atheism, or lost boyfriends/girlfriends, or were terrified they might go to Hell. And then they found r/atheism, and saw people making fun of religion, and insulting it, in really REALLY offensive ways. And no one was striking them down with lightning. No one was shouting them down. No one was doing much of anything at all. And to see this taboo violated in the most shocking possible way with no repercussions sort of broke the spell for them, like as long as people were behaving respectfully to religion, even respectfully disagreeing, it still had this aura of invincibility about it, but if some perfectly normal person can post a a stupid comic where Jesus has gay sex with Mohammed, then there's this whole other world out there where religion holds no power.

Gilbert tells the story of how when, as a young Christian struggling with doubt, he would read r/atheism to remind himself that atheists could be pretty awful. r/atheism is doing a bad job at being the sort of people who can convert Gilbert, and the new mods' policy of "you should have more civil and intellectual discussions" might work better on him. I think it would work better on me too.

But there is – previously unappreciated by me – a large population of people for whom really dumb offensive strawmannish memes are exactly what they need.

III.

A friend described his experiences in the Landmark Forum's self-improvement workshop. He said their modus operandi was to get people to take responsibility for the outcome of their actions. His example was an office worker who always did substandard work, and was always making excuses like "My boss doesn't s support me" or "My computer system isn't good enough" or "My coworkers aren't pulling their fair share." Landmark says those kinds of excuses are what's keeping you back. And they taught (again, according to this one person) that the solution was to treat everything that happens in your life as your responsiblity – no excuses, just "it was my fault" or "it's to my credit".

Then a few days later, I was reading a book on therapy which contained the phrase (I copied it down to make sure I got it right) "Don't be so hard on yourself. No one else is as hard on yourself as you are. You are your own worst critic."

Notice that this encodes the exact opposite assumption. Landmark claims its members are biased against ever thinking ill of themselves, even when they deserve it. The therapy book claims that patients are biased towards always thinking ill of themselves, even when they don't deserve it.

And you know, both claims are probably spot on. There are definitely people who are too hard on themselves. Ozy Frantz has done an amazing job of getting me and many other people inclined towards skepticism about feminist and transgender issues, engaging with us, and gradually convincing us to be more respectful and aware through sheer kindness and willingness to engage people reasonably on every part of the political spectrum. Two days ago some people on Twitter – who were angry Ozy said one need not boycott everything Orson Scott Card has ever written just because he's against gay marriage – told Ozy they weren't a real transgender person and suggested lots of people secretly disliked them. And instead of doing what I would do and telling the trolls to go to hell, Ozy freaked out and worried they was doing everything wrong and decided to delete everything they had ever written online. I know Ozy is their own worst critic and if that therapy book was aimed at people like them, it was entirely correct to say what it said.

On the other hand, I look at people like Amy's Baking Company, who are obviously terrible people, who get a high-status professional chef as well as thousands of random joes informing them of exactly what they are doing wrong, who are so clearly in the wrong that it seems impossible not to realize it – and who then go on to attribute the negativity to a "conspiracy" against them and deny any wrongdoing. They could probably use some Landmark.

IV.

In a recent essay I complained about bravery debates, arguments where people boast about how brave they are to take an unorthodox and persecuted position, and their opponents counter that they're not persecuted heretics, they're a vast leviathan persecuting everyone else. But I think I underestimated an important reason why some debates have to be bravery debates.

Suppose there are two sides to an issue. Be more or less selfish. Post more or less offensive atheist memes. Be more or less willing to blame and criticize yourself.

There are some people who need to hear each side of the issue. Some people really need to hear the advice "It's okay to be selfish sometimes!" Other people really need

to hear the advice “You are being way too selfish and it’s not okay.”

It’s really hard to target advice at exactly the people who need it. You can’t go around giving everyone surveys to see how selfish they are, and give half of them Atlas Shrugged and half of them the collected works of Peter Singer. You can’t even write really complicated books on how to tell whether you need more or less selfishness in your life – they’re not going to be as buyable, as readable, or as memorable as Atlas Shrugged. To a first approximation, all you can do is saturate society with pro-selfishness or anti-selfishness messages, and realize you’ll be hurting a select few people while helping the majority.

But in this case, it makes a really big deal what the majority actually is.

Suppose an Objectivist argues “Our culture has become too self-sacrificing! Everyone is told their entire life that the only purpose of living is to work for other people. As a result, people are miserable and no one is allowed to enjoy themselves at all.” If they’re right, then helping spread Objectivism is probably a good idea – it will help these legions of poor insufficiently-selfish people, but there will be very few too-selfish-already people who will be screwed up by the advice.

But suppose Peter Singer argues “We live in a culture of selfishness! Everyone is always told to look out for number one, and the poor are completely neglected!” Well, then we want to give everyone the collected works of Peter Singer so we can solve this problem, and we don’t have to worry about accidentally traumatizing the poor self-sacrificing people more, because we’ve already agreed there aren’t very many of these at all.

It’s much easier to be charitable in political debates when you view the two participants as coming from two different cultures that err on opposite sides, each trying to propose advice that would help their own culture, each being tragically unaware that the other culture exists.

A lot of the time this happens when one person is from a dysfunctional community and suggesting very strong measures against some problem the community faces, and the other person is from a functional community and thinks the first person is being extreme, fanatical or persecutory.

This happens a lot among, once again, atheists. One guy is like “WE NEED TO DESTROY RELIGION IT CORRUPTS EVERYTHING IT TOUCHES ANYONE WHO MAKES ANY COMPROMISES WITH IT IS A TRAITOR KILL KILL KILL.” And the other guy is like “Hello? Religion may not be literally true, but it usually just makes people feel more comfortable and inspires them to do nice things and we don’t want to look like huge jerks here.” Usually the first guy was raised Jehovah’s Witness and the second guy was raised Moralistic Therapeutic Deist.

But I’ve also sometimes had this issue when I talk to feminists. They’re like “Guys need to be more concerned about women’s boundaries, and women need to be willing to shame and embarrass guys who hit on them inappropriately.” And maybe they spent high school hanging out with bros on the football team who thought asking women’s consent was a boring technicality, and I spent high school hanging out entirely with extremely considerate but very shy geeks who spent their teenage years in a state of nightmarish loneliness and depression because they were too scared to ask out women because the woman might try to shame and embarrass them for it.

And the big one is trust. There are so many people from extremely functional communities saying that people need to be more trusting and kind and take people at their word more often, and so many people from dysfunctional communities saying that's not how it works. Both are no doubt backed by ample advice from their own lives.

A blog like this one probably should promote the opinions and advice most likely to be underrepresented in the blog-reading populace (which is totally different from the populace at large). But this might convince "thought leaders", who then use it to inspire change in the populace at large, which will probably be in the wrong direction. I think most of my friends are too leftist but society as a whole is too rightist – should I spread leftist or rightist memes among my friends?

I feel pretty okay about both being sort of a libertarian and writing an essay arguing against libertarianism, because the world generally isn't libertarian enough but the sorts of people who read long online political essays generally are way more libertarian than can possibly be healthy.

The Virtue of Silence

Leah Libresco writes a couple of essays (1, 2) on an ethical dilemma reported in the New York Times. In the course of a confidential medical history, a doctor hears her patient is suffering from stress-related complaints after having sent an innocent man to prison. The doctor wants to know whether it is ethical to report the matter to the police. The Times' columnist says yes – it would save the poor prisoner. Leah says no – violating medical confidentiality creates an expectation that medical confidentiality will be violated in the future, thus dooming patients who are too afraid to talk about drug use or gay sex or other potentially embarrassing but important medical risk factors.

But both sides are ignoring the much bigger dilemma lurking one meta-level up: is it ethical to debate this dilemma in the *New York Times*?

Let's look more closely at that phrase "violating medical confidentiality creates an expectation that medical confidentiality will be violated in the future." There's a very abstruse angels-and-clockwork interpretation of "creates an expectation" where, by making the decision to violate confidentiality, you are altering the Platonic machinery of the Universe in a way that allows other beings who know your source code to determine that you will do this. But most people don't have the decision theory to understand this, and anyway most doctors do not publish their source code online.

The way "creates an expectation" pans out in *our* universe is that somebody hears that a doctor violated medical confidentiality, and that person tells someone else, and that person tells someone else, until eventually someone who was going to tell their doctor about having gay sex with drugs remembers having heard the story and decides not to.

How exactly would people hear about this doctor who revealed the innocence of the prisoner? Through the ensuing court case? Nah. Most people wouldn't obsessively read the minutes of every single case at the local courthouse unless of course it has something to do with gender. Really, the only way that someone could hear about a doctor violating medical confidentiality is if she, like, somehow got a description of her intention to do so published in meticulous detail in the *New York Times*.

Oh, *right*.

The entire negative effect of the doctor breaking her promise is that it would make people doubt medical confidentiality in the future. But *whether or not the doctor ends up breaking her promise*, thousands of New York Times readers now know that doctors strongly consider breaking medical confidentiality, and that ethics columnists tell them it's okay to do so. It seems like the whether the doctor actually keeps her promise or not in this particular case is of miniscule importance compared to the damage that the column has already done.

Silence is a *hard* virtue. All the other virtues have the advantage that, when you practice them, people will praise you. Sometimes if your moral system is very different from your friends' people will attack you for your virtues, but getting attacked by sufficiently horrible people can sometimes be just as gratifying as praise. But if you stay silent, there's no praise *and* no attacks. By definition, no one even knows you made a courageous moral choice.

(Eliezer mentioned in the comments of my Against Bravery Debates that he's spent a couple decades pushing ideas almost everyone else thinks are crackpot, and he's never appealed to bravery at all. He is one hundred percent correct and I have one hundred percent never noticed despite reading almost everything he's written for several years. That's the Virtue of Silence for you.)

(I had like five much better examples here, all of which would be very clever, and each time I had to catch myself and say "Wait a second, by bringing that up I'm violating the virtue I'm supposed to be pushing here, aren't I?")

One example of silence I deeply appreciate is *people who don't talk about the latest viral issue*. I'm trying to think of an example that's not too destructive to bring up... hmmmm...go for something old...Elevatorgate! Nearly everyone who talked about Elevatorgate mentioned that it was outrageous that the blogosphere was making such a big deal about it, missing the similarity to the old adage that "you aren't stuck in traffic, you *are* traffic." Somewhere there was someone who wanted to write about Elevatorgate, thought about it, and *decided not to*. That person deserves the sincere thanks of a grateful Internet.

So having made the case for the other side of the confidentiality-newspaper meta-dilemma, am I *actually* pushing the claim that it is a moral law not to publicize information that could have bad consequences?

But I notice that this sort of thing almost always ends up making people angry and having a perverse effect where demands not to draw Mohammed turn into Everyone Draw Mohammed Day (see: Streisand Effect). It also sometimes snowballs to the point where not only can you not talk about X, but you can't talk about the demand not to talk about X because that would be referring to X obliquely, and you can't talk about the demand not to talk about the demand to talk about X, until eventually you climb up so many meta-levels that you collapse from hypoxia and have to be rescued by Sherpas. Then you get a "callout culture" where people try to gain easy Virtue Points by telling people discussing issues that they Should Not Be Discussing Them and other people try to gain easy Virtue Points by being the Brave Defender of Freedom of Speech.

And maybe that's useful if it's something like gender where everyone wants to talk about it all the time anyway, but we don't really need to do that to medical confidentiality, do we?

Maybe this is one of those rare cases where the word "supererogatory" might be useful. Yelling at people who talk about violations of medical confidentiality would just lead to "ARE OUR DISCUSSIONS OF MEDICAL CONFIDENTIALITY BEING SILENCED??!?" on the front page of the *New York Times*. And fretting over talking about it with your friends, or publishing a blog article about it (cough) is probably on the moral level of those Jains who walk everywhere with a broom in front of them so that they don't accidentally squash any bugs. But if someone is really really concerned about it and wants to be a great person, then yeah, I think writing to the *New York Times* about it requires a bit of thought.

And since I *am* publishing a blog article about it (VIRTUE OF SILENCE IS REALLY HARD!) let me restore some Virtue Points by confirming that I will *not* betray private patient information of this sort if such a dilemma comes up except when legally required. Trust me, I'm a doctor.

Proving Too Much

The fallacy of Proving Too Much is when you challenge an argument because, in addition to proving its intended conclusion, it also proves obviously false conclusions. For example, if someone says “You can’t be an atheist, because it’s impossible to disprove the existence of God”, you can answer “That argument proves too much. If we accept it, we must also accept that you can’t disbelieve in Bigfoot, since it’s impossible to disprove his existence as well.”

I love this tactic *so much*. I only learned it had a name quite recently, but it’s been my default style of argument for years. It neatly cuts through complicated issues that might otherwise be totally irresolvable.

Because here is a fundamental principle of the Dark Arts – you don’t need an argument that can’t be disproven, only an argument that can’t be disproven in the amount of time your opponent has available.

In a presidential debate, where your opponent has three minutes, that means all you need to do is come up with an argument whose disproof is inferentially distant enough from your audience that it will take your opponent more than three minutes to explain it, or your audience more than three minutes’ worth of mental effort to understand the explanation.

The noncentral fallacy is the easiest way to do this. “Martin Luther King was a criminal!” “Although what you say is technically correct, categories don’t work in the way your statement is impl – ” “Oh, sorry, time’s up.”

But pretty much anything that assumes a classical Aristotelian view of concepts/objects is gold here. The same is true of any deontological rules your audience might be attached to.

I tend to get stuck in the position of having argue against those Dark Artsy tactics pretty often. And the great thing about Proving Too Much is that it can demolish an entire complicated argument based on all sorts of hard-to-tease-apart axioms in a split second. For example, *After Virtue* gave (though it does not endorse) this example of deontological reasoning:

> I cannot will that my mother should have had an abortion when she was pregnant with me, except perhaps if it had been certain that the embryo was dead or gravely damaged. But if I cannot will this in my own case, how can I consistently deny to others the right to life that I claim for myself? I would break the so-called Golden Rule unless I denied that a mother in general has a right to an abortion.

It seemed unfair for me to move on in the book without at least checking whether this argument was correct and I should re-evaluate my pro-choice position. But that would require sorting through all the weird baggage here, like what it means to will something, and whether your obligations to potential people are the same as your obligations to real people, and how to apply the Golden Rule across different levels of potentiality.

Instead I just thought to myself: “Imagine my mother had raped my father, leading to my conception. I cannot will that a policeman had prevented this rape, but I also do

not want to enshrine the general principle that policemen in general have no right to prevent rape. Therefore, this argument proves too much." It took all of five seconds.

Sometimes a quick Proving Too Much can tear apart extremely subtle philosophical arguments that have been debated for centuries. For example, Pascal's Wager also proves Pascal's Mugging (they may both be correct, but bringing the Mugging in at least proves ignoring their correctness to be a reasonable and impossible-to-critique life choice). And Anselm's Ontological Argument seems much less foreboding when you realize it can double as a method for creating jelly donuts on demand.

Interestingly, I think that one of the examples of proving too much on Wikipedia can itself be demolished by a proving too much argument, but I'm not going to say which one it is because I want to see if other people independently come to the same conclusion.

Beware Isolated Demands For Rigor

I.

From Identity, Personal Identity, and the Self by John Perry:

> "There is something about practical things that knocks us off our philosophical high horses. Perhaps Heraclitus really thought he couldn't step in the same river twice. Perhaps he even received tenure for that contribution to philosophy. But suppose some other ancient had claimed to have as much right as Heraclitus did to an ox Heraclitus had bought, on the grounds that since the animal had changed, it wasn't the same one he had bought and so was up for grabs. Heraclitus would have quickly come up with some ersatz, watered-down version of identity of practical value for dealing with property rights, oxen, lyres, vineyards, and the like. And then he might have wondered if that watered-down vulgar sense of identity might be a considerably more valuable concept than a pure and philosophical sort of identity that nothing has.

Okay, but I can think of something worse than that.

Imagine Heraclitus as a cattle rustler in the Old West. Every time a rancher catches him at his nefarious business, he patiently explains to them that identity doesn't exist, and therefore the same argument against private property as made above. Flummoxed, they're unable to think of a response before he rides off into the sunset.

But then when Heraclitus himself needs the concept of stable personal identity for something – maybe he wants to deposit his ill-gotten gains in the bank with certainty that the banker will give it back to him next time he shows up to withdraw it, or maybe he wants to bribe the sheriff to ignore his activities for the next while – all of a sudden Heraclitus is willing to tolerate the watered-down vulgar sense of identity like everyone else.

(actually, I can think of something even worse than that, which is a TV western based on this premise, where a roving band of pre-Socratic desperadoes terrorizes Texas. The climax is no doubt when the hero strides onto Main Street, revolver in hand, saying "There's a new sheriff in town." And Parmenides gruffly responds "No, I'm pretty sure that's impossible.")

At its best, philosophy is a revolutionary pursuit that dissolves our common-sense intuitions and exposes the possibility of much deeper structures behind them. One can respond by becoming a saint or madman, or by becoming a pragmatist who is willing to continue to participate in human society while also understanding its theoretical limitations. Both are respectable career paths.

The problem is when someone chooses to apply philosophical rigor selectively.

Heraclitus could drown in his deeper understanding of personal identity and become a holy madman, eschewing material things and taking no care for the morrow because he does not believe there is any consistent self to experience it. Or he could engage with it from afar, becoming a wise scholar who participating in earthly affairs while drawing equanimity from the realization that there is a sense in which all his accomplishments will be impermanent.

But if he only applies his new theory when he wants other people's cows, then we have a problem. Philosophical rigor, usually a virtue, has been debased to an isolated demand for rigor in cases where it benefits Heraclitus.

A fair use of philosophical rigor would prevent both Heraclitus and his victims from owning property, and thus either collapse under its own impracticality or usher in a revolutionary new form of economic thinking. An isolated demand for philosophical rigor, applied by Heraclitus to other people but never the other way around, would merely give Heraclitus an unfair advantage in the existing system.

II.

A while ago I wrote a post called Military Strikes Are An Extremely Cheap Way To Help Foreigners which was a response to a Matt Yglesias post called the opposite. Yglesias was opposed to "humanitarian" military intervention (think the air strikes on ISIS going on right now, justified under the cause of preventing a genocide) and his argument was that this was extremely cost-ineffective compared to just giving the money to GiveWell's top-rated charity – at the time he was writing, malaria prevention.

I argued he was wrong about his numbers. But I also argued he was unfairly making an isolated demand for philosophical rigor.

Once you learn about utilitarianism and effective charity, you can become the holy madman, donating every cent you have beyond what is strictly necessary to survive and hold down a job to whatever the top rated charity is.

Or you can become the worldly scholar, continuing to fritter away your money on things like "hot water" and "food other than gruel" but appreciating the effective-utilitarian perspective and trying to make a few particularly important concessions to it.

Or you can use it to steal other people's cows. This is what I accused Matt Yglesias of doing. Presumably there are lots of government programs Yglesias supports – I suggested PBS – and he would never dream of demanding that we defund them in the hopes of donating the money to malaria prevention. But if for political reasons he doesn't support air strikes, suddenly that plan has to justify itself according to rigorous criteria that no government program that exists could possibly pass.

Government spending seems to be a particularly fertile case for this problem. I remember hearing some conservatives complain: sex education in public schools is an outrage, because my tax dollars are going to support something I believe is morally wrong.

This is, I guess, a demand for ethical rigor. That no one should ever be forced to pay for something they don't like. Apply it consistently, and conservatives shouldn't have to pay for sex ed, liberals shouldn't have to pay for wars, and libertarians shouldn't have to pay for anything, except maybe a $9.99 tax bill yearly to support the police and a minimal court system.

Applied consistently, you become the holy madman demanding either total anarchy or some kind of weird system of tax earmarks which would actually be pretty fun to think about. Or the worldly scholar with a strong appreciation for libertarian ideas who needs a really strong foundational justification for spending government money on things that a lot of people oppose.

Applied inconsistently, you're just stealing cows again, coming up with a clever argument against the programs you don't like while defending the ones you do.

III.

But this is the sort of uncouth behavior we expect of political partisans. What about science?

Suppose there are scientists on both sides of a controversial issue – for example, economists studying the minimum wage. One team that supports a minimum wage comes up with a pretty good study showing with $p < 0.05$ that minimum wages help the economy in some relevant way. The Science Czar (of course we have a science czar! We're not monsters!) notes that $p < 0.05$ is really a shoddy criterion that can prove anything and they should come back when they have $p < 0.01$. I have a huge amount of sympathy with the Science Czar on this one, by the way.

Soooo the team of economists spends another five years doing another study and finds with $p < 0.01$ that the minimum wage helps the economy in some important way. The Science Czar notes that their study was correlational only, and that correlational studies suck. We really can't show that minimum wages are any good without a randomized controlled trial. Luckily, the governments of every country in the world are totally game for splitting their countries in half and instituting different economic regimes in each part for ten years, so after a decade it comes out that in the randomized controlled trial the minimum wage helped the economy with $p < 0.01$. The Science Czar worries about publication bias. What if there were a lot of other teams who got all the countries in the world to split in half and institute different wage policies in each of the two territories for one decade, but they weren't published because their results weren't interesting enough? Everything the Science Czar has said so far makes perfect sense and he is to be commended for his rigor and commitment to the job. Science is really hard and even tiny methodological mistakes can in principle invalidate an entire field.

But now suppose that a team shows that, in a sample of six restaurants in Podunk Ohio, there was a nonsignificant trend towards the minimum wage making things a little worse.

And the Science Czar says: awesome! That solves that debate, minimum wage is bad, let's move on to investigating nominal GDP targeting.

Now it looks like the Science Czar is just a jerk who's really against minimum wage. All his knowledge of the standards of scientific rigor are going not towards bettering science, but toward worsering science. He's not trying to create a revolutionary new scientific regime, he's taking pot shots.

I see this a lot in medicine. Someone jumps on a new study showing the selenium or chromium or plutonium or whatever cures cancer. It is brought up that no, really, the medical community has investigated this sort of thing before, and it has always been found that it doesn't.

"Well, maybe the medical community wasn't investigating it the right way! Maybe the investigators were biased! Maybe they didn't randomize right! Maybe they used a population unusually susceptible to cancer-getting! Ninety percent of medical studies are wrong! Those twenty experiments showing a lack of effect could be total bunk!"

Yes, maybe these things happened in each of the twenty studies that disagree with you.

Or maybe they happened in the one contrarian study you are getting so excited about.

IV.

The unholy combination of isolated demands for philosophical rigor and isolated demands for scientific rigor is isolated demands for mathematical-statistical-conceptual rigor, ie the sort of thing this blog has been talking about all week.

I have already been made fun of for how many different things I am metaphorically comparing IQ to – speed, blood pressure, comas – so I guess it can't hurt to add another example I only thought of today. How about crime? It's usually measured by crime rate – a made-up statistic that combines subfactors like arson (maybe higher when fire insurance pays out better), property damage (maybe higher during periods of ethnic tension and frequent riots) and theft (maybe higher when income inequality is worse). There is assumed to be a General Factor Of Crime (presumably caused by things like poor policing, dark alleys, broken families, et cetera) but I would be extremely surprised if anyone had ever proven Beyond A Shadow Of A Doubt that the factor analysis works out here.

When Cosma Shalizi says he's not sure about the factor analysis in IQ, I have no quarrel with him, because Cosma Shalizi's response to everything in the world is to glare at it for not being sufficiently statistically rigorous.

But when other people are totally happy to talk about speed and blood pressure and comas and the crime rate, and then suddenly switch to a position that we can't talk about IQ at all unless we have a perfect factor-analytical proof of its obeying certain statistical rules, then I worry they're just out to steal cows.

Likewise, if someone were to just never acknowledge any sorts of groups of objects except those that could be statistically proven to fall out into absolutely separate clusters in which variance within each cluster is less than variance between clusters, well, at least they would be fun to talk to at dinner parties.

But when people never even begin to question the idea of different cultures but make exacting demands of anyone before they can talk about different races – even though the two ideas are statistically isomorphic – then I think they're just out to steal cows.

So this is another technique for avoiding Eulering – is your interlocutor equally willing to apply their complex mathematical argument to everything else.

I think if I hadn't known anything about Bayesian probability, I would have examined the McGrews' Bayesian argument for the Gospels by seeing if it applied equally well to Mormonism, the control group for Christianity.

V.

The old man stamped his boot in the red dirt, kicking up a tiny cloud of dust. "There's a new sheriff in town," he told them.

"No, I'm pretty sure that's impossible," says Parmenides. "There's no such thing as change, only the appearance thereof."

“Well then,” says the old man, “I reckon you won’t mind the false illusion of your surroundings appearing to change into a jail cell.” And he took out his six-shooter and held it steady.

“Hold on,” said Thales. “We don’t want any trouble here. All is water, so all we did was steal a little bit of water from people. We can give you some water back, and everything will be even, right?” He gestured to a watering trough for horses on the side of the street, which was full of the stuff.

“Just so long as you don’t mind being sprayed with some very hard water from my squirt gun,” the old man answered, and the six-shooter was pointed at the Milesian now.

“Ha!” said Zeno of Elea. “You don’t scare us. In order to hit Thales, your bullet would have to get halfway to him, then half of the remaining distance, and so on. But that would require an infinite number of steps, therefore it is impossible.”

“Sorry,” said the old man, “I couldn’t hear you because it’s logically impossible for the sound waves encoding your speech to reach my ears.”

“We’re not even the same people as the guys who stole those cattle!” said Heraclitus. “Personal identity is an illusion!”

“Then you won’t mind coming to the courthouse with me,” replied the old man “to help the judge imprison some other people who look just like you.”

The last of them, the tall one, said nothing. He just raised his revolver in a fluid motion and shot at the old man.

The old man saw it coming and jumped out of the way. The air was briefly full of bullets. Bang! Thales went down! Bang bang! Heraclitus! Bang bang! Parmenides and Zeno. Bang bang bang! The old man was hit in the arm, but still standing. Bang bang bang bang…

It was just the old man and the tall one now. The tall one picked up his gun and fired. Nothing happened. Out of bullets.

The old man smiled wryly, his six-shooter still in his hand.

“I know what you’re thinking. You’re thinking – did he fire six shots, or only five? Well, you’ve got to ask yourself a question – do you feel lucky? Well, do you, punk?”

The tall one didn’t budge. “Man is the measure of all things,” said Protagoras. “If I believe you fired six shots, then by my personal epistemic standards, you fired six shots.”

The old man didn’t say anything.

“You see,” the Sophist continued. “Out of all of them, I alone was truly consistent. They all came up with clever theories, then abandoned them whenever it conflicted with their self-interest. I was more honest. I just said at the beginning that my self-interest determined truth, and so never suffered any temptation to depart from my position.”

The old man took off the bandana covering his face. “Man may be the measure of all things. But I’ve taken your measure, Protagoras, and found it wanting.”

“Socrates?!” the Sophist gasped.

“The only truly consistent people are the dead, Protagoras,” he said – and squeezed the trigger.

Transhumanist Fables

Once upon a time there were three little pigs who went out into the world to build their houses. The first pig was very lazy and built his house out of straw. The second pig was a little harder-working and built his house out of sticks. The third pig was the hardest-working of all, and built his house out of bricks. Then came the Big Bad Wolf. When he saw the house of straw, he huffed and he puffed and he blew the house down, eating the first little pig. When he saw the house of sticks, he huffed and he puffed and he blew the house down, eating the second little pig. When he saw the house of bricks, he got out a bazooka and blew the house to pieces, eating the third little pig.

Moral: Reality doesn't grade on a curve.

Once upon a time there was a big strong troll who lived under a bridge. A little goat went across the bridge, and the troll reached out to grab and eat the goat. "Wait, Mr. Troll!", the goat cried. "Soon my brother is coming, and he is even bigger than I am!" The troll let the goat pass, and soon came another goat, twice as big as the first. The troll reached out to grab and eat him, but the brother likewise objected, saying *his* brother was even bigger. Sure enough, a third goat arrived at the bridge, twice as big as the second, and the troll, now ready for a very hearty dinner, reached out to grab and eat him. "Wait!" said the third goat. "My brother is the biggest of us all!". So the troll let the third goat pass. Then came the fourth goat, who was hundreds of miles tall and blotted out the sun, whose very steps caused earthquakes and made the rivers change course. Without even noticing, he stepped on bridge and troll, pulverizing both to bits.

Moral: Sometimes growth is superexponential.

Once upon a time, Chicken Little ran to her friend Henny Penny. "The sky is falling!" she shouted. "We must tell the king!" Henny Penny joined her, and together they headed toward the capital. On their way they run into their friend Goosey Loosey. "The sky is falling!" they shouted. "We must tell the king!" Goosey Loosey joined them, and together they headed toward the capital. On their way, they ran into the cunning Foxy Loxy. "The sky is falling!" they shouted. "We must tell the king!" "Oh," said Foxy Loxy. "I know a shortcut to the palace. Follow me into my den." So the birds all followed Foxy Loxy into his den, where he ate them all, laughing all the while about how gullible they were. Then an asteroid hit Earth, killing everyone.

Moral: Beware the absurdity heuristic.

Once upon a time, a young boy named Jack lived with his mother. Their family was very poor and owned only a single cow. "Go sell this cow at the market," Jack's mother told him, "so we will have food to eat for the winter." Jack went to the market and came back with three beans. "These are magic beans!" he told his mother. "A man told me that when we plant them, they will grow into a beanstalk leading to a land of infinite riches." His mother pooh – poohed him and threw the beans in the ground angrily. That winter, they both died of hunger.

Moral: Good decision theories should be able to resist Pascal's Mugging.

Once upon a time, there was an old woodcutter who had no son. He made a little marionette out of pine wood and named it Pinocchio. Then he wished upon a star that it could become a real boy. The star turned out to be the evil Red Fairy, who brought Pinocchio to life, but told him that if he wanted to be a real boy he must murder everyone in the village. That night, Pinocchio took his father's saw and killed Gepetto and everyone else in town.

Moral: Never create an intelligence unless you are certain it will share your values.

Once upon a time, an evil witch transformed a prince into a frog, telling him that only the kiss of a princess could restore him to his proper form. But although he searched around the world, he could find no princess who was willing to kiss a hideous little frog. Finally, he went to the Wise Wizard. "Gender is a social construct," said the Wise Wizard. "Just declare your gender identity to be female, then kiss yourself on the hand or something." So the frog did that, returned to human form, and ruled the land for many years as a wise and benevolent queen.

Moral: Ability to self-modify is just *ridiculously* powerful.

…And I Show You How Deep The Rabbit Hole Goes

.

Seen on Tumblr, along with associated discussion:

Yellow:

People's minds are heartbreaking. Not because people are so bad, but because they're so good.

Nobody is the villain of their own life story. You must have read hundreds of minds by now, and it's true. Everybody thinks of themselves as an honest guy or gal just trying to get by, constantly under assault by circumstances and The System and hundreds and hundreds of assholes. They don't just sort of believe this. They really believe it. You almost believe it yourself, when you're deep into a reading. You can very clearly see the structure of evidence they've built up to support their narrative, and even though it looks silly to you, you can see why they will never escape it from the inside. You can see how every insult, every failure, no matter how deserved, is a totally unexpected kick in the gut.

When you chose the yellow pill, you had high hopes of becoming a spy, or a gossip columnist, or just the world's greatest saleswoman. The thought of doing any of those things sickens you now. There is too much anguish in the world already. You feel like any of those things would be a violation. You briefly try to become a therapist, but it turns out that actually knowing everything about your client's mind is horrendously countertherapeutic. Freud can say whatever he wants against defense mechanisms, but without them, you're defenseless. Your sessions are spent in incisive cutting into your clients' deepest insecurities alternating with desperate reassurance that they are good people anyway.

Also, men. You knew, in a vague way, that men thought about sex all the time. But you didn't realize the, um, content of some of their sexual fantasies. Is it even *legal* to fantasize about that? You want to be disgusted with them. But you realize that if you were as horny as they were all the time, you'd do much the same.

You give up. You become a forest ranger. Not the type who helps people explore the forest. The other type. The type where you hang out in a small cabin in the middle of the mountains and never talk to anybody. The only living thing you encounter is the occasional bear. It always thinks that it is a good bear, a proper bear, that a bear-hating world has it out for them in particular. You do nothing to disabuse it of this notion.

Green

The first thing you do after taking the green pill is become a sparrow. You soar across the landscape, feeling truly free for the first time in your life.

You make it about five minutes before a hawk swoops down and grabs you. Turns out there's an excellent reason real sparrows don't soar freely across the open sky all day. Moments before your bones are ground in two by its fierce beak, you turn back into a human. You fall like a stone. You need to turn into a sparrow again, but the hawk is still there, grabbing on to one of your legs, refusing to let go of its prize just because of this momentary setback. You frantically wave your arms and shout at it, trying to scare it away. Finally it flaps away, feeling cheated, and you become a sparrow again just in time to give yourself a relatively soft landing.

After a few weeks of downtime while you wait for your leg to recover, you become a fish. This time you're smarter. You become a great white shark, apex of the food chain. You will explore the wonders of the ocean depths within the body of an invincible killing machine.

Well, long story short, it is totally unfair that colossal cannibal great white sharks were a thing and if you had known this was the way Nature worked you never would have gone along with this green pill business.

You escape by turning into a blue whale. Nothing eats blue whales, right? You remember that from your biology class. It is definitely true.

The last thing you hear is somebody shouting "We found one!" in Japanese. The last thing you feel is a harpoon piercing your skull. Everything goes black.

Blue

Okay, so you see Florence and Jerusalem and Kyoto in an action-packed afternoon. You teleport to the top of Everest because it is there, then go to the bottom of the Marianas Trench. You visit the Amazon Rainforest, the Sahara Desert, and the South Pole. It takes about a week before you've exhausted all of the interesting tourist sites. Now what?

You go to the Moon, then Mars, then Titan. These turn out to be even more boring. Once you get over the exhilaration of being on Mars, there's not a lot to do except look at rocks. You wonder how the Curiosity Rover lasted so long without dying of boredom.

You go further afield. Alpha Centauri A has five planets orbiting it. The second one is covered with water. You don't see anything that looks alive in the ocean, though. The fourth has a big gash in it, like it almost split in two. The fifth has weird stalactite-like mountains.

What would be really interesting would be another planet with life, even intelligent life. You teleport further and further afield. Tau Ceti. Epsilon Eridani. The galactic core. You see enough geology to give scientists back on Earth excitement-induced seizures for the nest hundred years, if only you were to tell them about it, which you don't. But nothing alive. Not so much as a sea cucumber.

You head back to Earth less and less frequently now. Starvation is a physical danger, so it doesn't bother you, though every so often you do like to relax and eat a nice warm meal. But then it's back to work. You start to think the Milky Way is a dead zone. What about Andromeda…?

Orange

You never really realized how incompetent everyone else was, or how much it annoys you.

You were a consultant, a good one, but you felt like mastering all human skills would make you better. So you took the orange pill. The next day you go in to advise a tech company on how they manage the programmers, and you realize that not only are they managing the programmers badly, but the programmers aren't even writing code very well. You could write their system in half the time. The layout of their office is entirely out of sync with the best-studied ergonomic principles. And the Chinese translation of their user manual makes several basic errors that anybody with an encyclopaedic knowledge of relative clauses in Mandarin should have been able to figure out.

You once read about something called Gell-Mann Amnesia, where physicists notice that everything the mainstream says about physics is laughably wrong but think the rest is okay, doctors notice that everything the mainstream says about medicine is laughably wrong but think the rest is okay, et cetera. You do not have Gell-Mann Amnesia. Everyone is terrible at everything all the time, and it pisses you off.

You gain a reputation both for brilliance and for fearsomeness. Everybody respects you, but nobody wants to hire you. You bounce from industry to industry, usually doing jobs for the people at the top whose jobs are so important that the need to get them done right overrides their desire to avoid contact with you.

One year you get an offer you can't refuse from the King of Saudi Arabia. He's worried about sedition in the royal family, and wants your advice as a consultant for how to ensure his government is stable. You travel to Riyadh, and find that the entire country is a mess. His security forces are idiots. But the King is also an idiot, and refuses to believe you or listen to your recommendations. He tells you things can't possibly be as bad as all that. You tell him you'll prove that they are.

You didn't *plan* to become the King of Saudi Arabia, per se. It just sort of happened when your demonstration of how rebels in the military might launch a coup went better than you expected. Sometimes you forget how incompetent everybody else is. You need to keep reminding yourself of that. But not right now. Right now you're busy building your new capital. How come nobody else is any good at urban planning?

Red

You choose the red pill. BRUTE STRENGTH! That's what's important and valuable in this twenty-first-century economy, right? Some people tell you it isn't, but they don't seem to have a lot of BRUTE STRENGTH, so what do they know?

You become a weightlifter. Able to lift thousands of pounds with a single hand, you easily overpower the competition and are crowned whatever the heck it is you get crowned when you WIN WEIGHTLIFTING CONTESTS. But this fails to translate into lucrative endorsement contracts. Nobody wants their spokesman to be a bodybuilder without a sixpack, and although you used to be pretty buff, you're getting scrawnier by the day. Your personal trainer tells you that you only maintain muscle mass by doing difficult work at the limit of your ability, but your abilities don't seem to *have* any limits. Everything is so easy for you that your body just shrugs it off effortlessly. Somehow your BRUTE STRENGTH failed to anticipate this possibility. If only there was a way to solve your problem by BEING VERY STRONG.

Maybe the Internet can help. You Google “red pill advice”. The sites you get don’t seem to bear on your specific problem, exactly, but they are VERY FASCINATING. You learn lots of surprising things about gender roles that you didn’t know before. It seems that women like men who have BRUTE STRENGTH. This is relevant to your interests!

You leave the bodybuilding circuit behind and start frequenting nightclubs, where you constantly boast of your BRUTE STRENGTH to PROVE HOW ALPHA YOU ARE. A lot of people seem kind of creeped out by a scrawny guy with no muscles going up to every woman he sees and boasting of his BRUTE STRENGTH, but the Internet tells you that is because they are BETA CUCKOLD ORBITERS.

Somebody told you once that Internet sites are sometimes inaccurate. You hope it’s not true. How could you figure out which are the inaccurate ones using BRUTE STRENGTH?

Pink

You were always pretty, but never *pretty* pretty. A couple of guys liked you, but they were never the ones you were into. It was all crushingly unfair. So you took the pink pill, so that no one would ever be able to not love you again.

You find Tyler. Tyler is a hunk. He’d never shown any interest in you before, no matter how much you flirted with him. You touch him on the arm. His eyes light up.

“Kiss me,” you say.

Tyler kisses you. Then he gets a weird look on his face. “Why am I kissing you?” he asks. “I’m sorry. I don’t know what came over me.” Then he walks off.

You wish you had thought further before accepting a superpower that makes people love you when you touch them, but goes away after you touch them a second time. Having people love you is a lot less sexy when you can’t touch them. You start to feel a deep sense of kinship with King Midas.

You stop dating. What’s the point? They’ll just stop liking you when you touch them a second time. You live alone with a bunch of cats who purr when you pet them, then hiss when you pet them again.

One night you’re in a bar drinking your sorrows away when a man comes up to your table. “Hey!” he says, “nice hair. Is it real? I’m the strongest person in the world.” He lifts your table over his head with one hand to demonstrate. You are immediately smitten by his BRUTE STRENGTH and ALPHA MALE BEHAVIOR. You *must* have him.

You touch his arm. His eyes light up. “Come back to my place,” you say. “But don’t touch me.”

He seems a little put out by this latter request, but the heat of his passion is so strong he would do anything you ask. You move in together and are married a few contact-free months later. Every so often you wonder what it would be like to stroke him, or feel his scrawny arm on your shoulder. But it doesn’t bother you much. You’re happy to just hang out, basking in how STRONG and ALPHA he is.

Grey

Technology! That's what's important and valuable in this twenty-first-century economy, right? Right! For example, ever since you took the grey pill, an increasingly large share of national GDP has come from ATMs giving you cash because you ask them to.

Your luck finally ends outside a bank in Kansas, when a whole squad of FBI agents ambushes you. You briefly consider going all Emperor Palpatine on their asses, but caution wins out and you allow yourself to be arrested.

Not wanting to end up on an autopsy table in Roswell, you explain that you're a perfectly ordinary master hacker. The government offers you a plea bargain: they'll drop charges if you help the military with cyber-security. You worry that your bluff has been called until you realize that, in fact, you *are* a master hacker. So you join the NSA and begin an illustrious career hacking into Russian databases, stalling Iranian centrifuges, and causing Chinese military systems to crash at inconvenient times. No one ever suspects you are anything more than very good at programming.

Once again, your luck runs out. Your handlers ask you to hack into the personal files of a mysterious new player on the world stage, a man named William who seems to have carved himself an empire in the Middle East. You don't find anything too damning, but you turn over what you've got.

A few days later, you're lying in bed drifting off to sleep when a man suddenly bursts in through your window brandishing a gun. Thinking quickly, you tell the gun to explode in his hands. Nothing happens. The man laughs. "It's a decoy gun," he said. "Just here to scare you. But you bother King William again, and next time I'm coming with a very real knife." He jumps back out of the window. You call the police, and of course the CIA and NSA get involved, but he is never caught.

After that, you're always looking over your shoulder. He *knew*. How did he know? The level of detective skills it would take in order to track you down and figure out your secret – it was astounding! Who *was* this King William?

You tell your handlers that you're no longer up for the job. They beg, cajole, threaten to reinstate your prison sentence, but you stand firm. Finally they transfer you to an easier assignment in the Moscow embassy. You make Vladimir Putin's phone start ringing at weird hours of the night so that he never gets enough sleep to think entirely clearly. It's an easy job, but rewarding, and no assassins ever bother you again.

Black

You know on an intellectual level that there are people who would choose something other than the black pill, just like you know on an intellectual level that there are people who shoot up schools. That doesn't mean you expect to ever *understand* it. You just wish you could have taken the black pill before you had to decide what pill to take, so that you could have analyzed your future conditional on taking each, and so made a more informed decision. But it's not like it was a very hard choice.

The basic principle is this – given a choice between A and B, you solemnly resolve to do A, then see what the future looks like. Then you solemnly resolve to do B, and do the same. By this method, you can determine the optimal choice in every situation, modulo the one month time horizon. You might not be able to decide what career to pursue, but you can sure as heck ace your job interview.

Also, a millisecond in the future is pretty indistinguishable from the present, so "seeing" a millisecond into the future gives you pretty much complete knowledge about the current state of the world.

You are so delighted by your omniscience and your ability to make near-optimal choices that it takes almost a year before you realize the true extent of your power.

You resolve, on the first day of every month, to write down what you see exactly a month ahead of you. But what you will see a month ahead of you is the piece of paper on which you have written down what you see a month ahead of *that*. In this manner, you can relay messages back to yourself from arbitrarily far into the future – at least up until your own death.

When you try this, you see yourself a month in the future, just finishing up writing a letter that reads as follows:

> Dear Past Self:
>
> In the year 2060, scientists invent an Immortality Serum. By this point we are of course fabulously wealthy, and we are one of the first people to partake of it. Combined with our ability to avoid accidents by looking into the future, this has allowed us to survive unexpectedly long.
>
> I am sending this from the year 963,445,028,777,216 AD. We are one of the last hundred people alive in the Universe. The sky is black and without stars; the inevitable progress of entropy has reduced almost all mass and energy to unusable heat. The Virgo Superconfederation, the main political unit at this stage of history, gathered the last few megatons of usable resources aboard this station so that at least one outpost of humanity could last long after all the planets had succumbed. The station has been fulfilling its purpose for about a billion years now, but we only have enough fuel left for another few weeks. After that, there's no more negentropy left anywhere in the universe except our own bodies. I have seen a month into the future. Nobody comes to save us.
>
> For the past several trillion years, our best scientists have been investigating how to reverse entropy and save the universe, or how to escape to a different universe in a lesser state of decay, or how to collect energy out of the waste heat which now fills the vast majority of the sky. All of these tasks have been proven impossible. There is no hope left, except for one thing.
>
> It's impossible to see the future, even if it's only a month ahead. Somehow, our black pill breaks the laws of physics. Despite having explored throughout the cosmos, my people have found no alien species, nor any signs that such species ever existed. Yet somebody made the black pill. If we understood that power, maybe we could use it to save reality from its inevitable decay.
>
> By sending this message back, I destroy my entire timeline. I do this in the hopes that you, in the carefree springtime of the universe, will be able to find the person who made these pills and escape doom in the way we could not.
>
> Yours truly,
> You From Almost A Quadrillion Years In The Future

ACT TWO

Red

You hit the punching bag. It bursts, sending punching-bag-filling spraying all over the room! You know that that would happen! It always happens when you hit a punching bag! Your wife gets really angry and tells you that we don't have enough money to be getting new punching bags all the time, but women hate it when you listen to what they say! The Internet told you that!

The doorbell rings. You tear the door off its hinges instead of opening it, just to show it who's boss. Standing on your porch is a man in black. He wears a black cloak, and his face is hidden by a black hood. He raises a weapon towards you.

This looks like one of the approximately 100% of problems that can be solved by BRUTE STRENGTH! You lunge at the man, but despite your super-speed, he steps out of the way easily, even gracefully, as if he had known you were going to do that all along. He squeezes the trigger. You jump out of the way, but it turns out to be more *into* the way, as he has shot exactly where you were jumping into. Something seems very odd about this. Your last conscious thought is that you wish you had enough BRUTE STRENGTH to figure out what is going on.

Pink

You come home from work to a living room full of punching-bag-parts. Your husband isn't home. You figure he knew you were going to chew him out for destroying another punching bag, and decided to make himself scarce. That lasts right up until you go into the kitchen and see a man dressed all in black, sitting at the table, as if he was expecting you.

You panic, then reach in to touch him. If he's an axe murderer or something, you'll seduce him, get him wrapped around your little finger, then order him to jump off a cliff to prove his love for you. It's nothing you haven't done before, though you don't like to think about it too much.

Except that this man has no bare skin anywhere. His robe covers his entire body, and even his hands are gloved. You try to reach in to touch his face, but he effortlessly manuevers away from you.

"I have your husband," he says, after you give up trying to enslave him with your magic. "He's alive and in a safe place."

"You're lying!" you answer. "He never would have surrendered to anyone! He's too alpha!"

The man nods. "I shot him with an elephant tranquilizer. He's locked up in a titanium cell underneath fifty feet of water. There's no way he can escape using BRUTE STRENGTH. If you ever want to see him again, you'll have to do what I say."

"Why? Why are you doing this to me?" you say, crying.

"I need the allegiance of some very special people," he said. "They won't listen to me just because I ask them to. But they might listen to me because *you* ask them to. I understand you are pretty special yourself. Help me get who I want, and when we are done here, I'll let you and your husband go."

There is ice in his voice. You shiver.

Grey

That night with the assassin was really scary. You swore you would never get involved in King William's business again. Why are you even considering this?

"Please?" she said, with her big puppy dog eyes.

Oh, right. Her. She's not even all that pretty. Well, pretty, but not *pretty* pretty. But somehow, when she touched you, it was like those movies where you hear a choir of angels singing in the background. You would do anything she said. You know you would.

"We need to know the layout of his palace compound," said the man in black. Was he with her? Were they dating? If they were dating, you'll kill him. It doesn't matter how creepy he is, you won't tolerate competition. But they're probably not dating. You notice how he flinches away from her, like he's afraid she might touch him.

"And it has to be me who helps?"

"I've, ah, simulated hundreds of different ways of getting access to the King. None of them hold much promise. His security is impeccable. Your special abilities are the only thing that can help us."

You sit down at your terminal. The Internet is slow; DC still doesn't have fiber optic. You've living here two years now, in a sort of retirement, ever since King William took over Russia and knocked the bottom out of the Putin-annoying business. William now controls the entire Old World, you hear, and is also Secretary-General of the United Nations and Pope of both the Catholic and the Coptic Churches. The United States is supposedly in a friendly coexistence with him, but you hear his supporters are gaining more and more power in Congress.

It only takes a few minutes' work before you have the documents you need. "He currently spends most of his time at the Rome compound," you say. "There are five different security systems. I can disable four of them. The last one is a complicated combination of electrical and mechanical that's not hooked into any computer system I'll be able to access. The only way to turn it off is from the control center, and the control center is on the *inside* of the perimeter."

The man in black nods, as if he'd been expecting that. "Come with me," he says. "We'll take care of it."

Blue

There are a hundred billion stars in the Milky Way. Each has an average of about one planet – some have many more, but a lot don't have planets at all.

If you can explore one planet every half-hour – and you can, it doesn't take too long to teleport to a planet, look around to see if there are plants and animals, and then move on to the next one – it would take you five million years to rule out life on every planet in the galaxy.

That's not practical. But, you think, life might spread. Life that originates on one planet might end up colonizing nearby planets and star systems. That means your best bet is to sample various regions of the galaxy, instead of going star by star.

That’s what you’ve been doing. You must have seen about a hundred thousand planets so far. Some of them have beggared your imagination. Whole worlds made entirely of amethyst. Planets with dozens of colorful moons that make the night sky look like a tree full of Christmas ornaments. Planets with black inky oceans or green copper mountains.

But no life. No life anywhere.

A few years ago, you felt yourself losing touch with your humanity. You made yourself promise that every year, you’d spend a week on Earth to remind yourself of the only world you’ve ever seen with a population. Now it seems like an unpleasant task, an annoying imposition. But then, that was why you made yourself promise. Because you knew that future-you wouldn’t do it unless they had to.

You teleport into a small Welsh hamlet. You’ve been away from other people so long, you might as well start small. No point going right into Times Square.

A person is standing right next to you. She reaches out her arm and touches you. You jump. How did she know you would –

“Hi,” she says.

You’re not a lesbian, but you can’t help noticing she is the most beautiful person you’ve ever seen, and you would do anything for her.

“I need your help.” A man dressed all in black is standing next to her.

“You should help him,” the most beautiful person you’ve ever seen tells you, and you immediately know you will do whatever he asks.

Orange

You are in your study working on a draft version of next year’s superweapon budget when you hear the door open. Four people you don’t recognize step into the room. A man dressed in black. Another man wearing a grey shirt, thick glasses and is that a *pocket protector?* A woman in pink, pretty but not *pretty* pretty. Another woman in blue, who stares through you, like her mind is somewhere else. All five of your security systems have been totally silent.

You press the button to call your bodyguards, but it’s not working. So you draw the gun out from under your desk and fire; you happen to be a master marksman, but the gun explodes in your face. You make a connection. A person from many years ago, who had the power to control all technology.

No time to think now. You’re on your feet; good thing you happen to be a black belt in every form of martial arts ever invented. The man in grey is trying to take out a weapon; you kick him in the gut before he can get it out, and he crumples over. You go for the woman in blue, but at the last second she teleports to the other side of the room. This *isn’t fair*.

You are about to go after the woman in pink, but something in her step, something in the position of the others makes you think they *want* you to attack her. You happen to be a master at reading microexpressions, so this is clear as day to you; you go after the man in black instead. He deftly sidesteps each of your attacks, almost as if he knows what you are going to do before you do it.

The woman in blue teleports behind you and kicks you in the back, hard. You fall over, and the woman in pink grabs your hand.

She is very, very beautiful. How did you miss that before? You feel a gush of horror that you almost punched such a beautiful face.

“We need your help,” she says.

You are too lovestruck to say anything.

“The pills,” said the man in black. “Can you make them?”

“No,” you say, truthfully. “Of course I tried. But I wouldn’t even know where to begin creating magic like that.”

“And you’ve mastered all human jobs and activities,” said the man in black. “Which means the pills weren’t created by any human.”

“But there aren’t any aliens,” said the woman in blue. “Not in this galaxy, at least. I’ve spent years looking. It’s totally dead.”

“It’s just as I thought,” said the man in black. He turns to you. “You’re the Pope now, right? Come with us. We’re going to need you to get a guy in northern Italy to give us something very important.”

Yellow

It is spring, now. Your favorite time in the forest. The snow has melted, the wildflowers have started to bloom, and the bears are coming out of hibernation. You’re walking down to the river when someone leaps out from behind a tree and touches you. You scream, then suddenly notice how beautiful she is.

Four other people shuffle out from behind the trees. You think one of them might be King William, the new world emperor, although that doesn’t really make sense.

“You’re probably wondering why I’ve called all of you together today…” said the man in black. You’re not actually wondering that, at least not in quite those terms, but the woman in pink seems be listening intently so you do the same in the hopes of impressing her.

“Somehow – and none of us can remember exactly how – each of us took a pill that gave us special powers. Mine was to see the future. I saw to the end of time, and received a message from the last people in the universe. They charged me with the task of finding the people who created these pills and asking them how entropy might be reversed.

But I couldn’t do it alone. I knew there were seven other people who had taken pills. One of us – Green – is dead. Another – Red – had nothing to contribute. The rest of us are here. With the help of Pink, Blue, and Gray, we’ve enlisted the help of Orange and his worldwide organization. Now we’re ready for the final stage of the plan. Yellow, you can read anybody’s mind from a picture, right?”

Yellow nods. “But it has to be a real photograph. I can’t just draw a stick figure and say it’s the President and read his mind. I tried that.”

Black is unfazed. “With the help of Orange, who among his many other accomplishments is the current Pope, I have obtained the Shroud of Turin. A perfect photographic representation of Jesus Christ, created by some unknown technology in the first century. And Jesus, I am told, is an incarnation of God.”

“As the current Pope, I suppose I would have to agree with that assessment,” says Orange. “Though as the current UN Secretary General, I am disturbed by your fanatical religious literalism.”

“Orange can do anything that humans can do, and says he can’t make the pills. Blue has searched the whole galaxy, and says there aren’t any aliens. That leaves only one suspect. God must have made these pills, which means He must know how to do it. If we can read His mind, we can steal his secrets.”

“As Pope,” says Orange, “I have to condemn this in the strongest possible terms. But as Lucasian Professor of Mathematics at Cambridge, I have to admit I’m intrigued by this opportunity to expand our knowledge.”

Black ignores him. “Yellow, will you do the honors?”

You want no part in this. “This is insane. Every time I read someone’s mind I regret it. Even if it’s a little kid or a bear or something. It’s too much for me. I can’t deal with all of their guilt and sorrow and broken dreams and everything. There is *no way* I am touching the mind of God Himself.”

“Pleeeeeease?” asks Pink, with big puppy dog eyes.

“Um,” you say.

“Don’t you know how this will go, anyway?” asks Blue. “Why don’t you just tell her what happens?”

“Um,” said Black. “This is actually the one thing I haven’t been able to see. I guess contact with God is inherently unpredictable, or something.”

“I have *such* a bad feeling about this,” you say.

“Pweeeeeeease?” says Pink. She actually says pweeeeeeease.

You sigh, take the shroud, and stare into the eyes of Weird Photographic Negative Jesus.

Black

It is the year 963,445,028,777,216 AD, and here you are in a space station orbiting the Galactic Core.

After handing Yellow the Shroud of Turin, the next thing you remember is waking up in a hospital bed. The doctor tells you that you’d been in a coma for the past forty one years.

Apparently Yellow went totally berserk after reading God’s mind. You don’t know the details and you don’t want to, but she immediately lashed out and used her superpowers to turn off the minds of everybody within radius, including both you and herself. You all went comatose, and probably would have starved to death in the middle of the forest if Orange’s supporters hadn’t launched a worldwide manhunt for

him. They took his body and the bodies of his friends back to Rome, where they were given the best possible medical care while a steward ruled over his empire.

After forty-one years of that, Yellow had a heart attack and died, breaking the spell and freeing the rest of you. Except Blue and Grey. They'd died as well. It was just you, Orange, and Pink now.

Oh, and Red. You'd hired a friend to watch over him in his titanium jail cell, and once it became clear you were never coming back, he'd had mercy and released the guy. Red had since made a meager living selling the world's worst body-building videos, which were so bad they had gained a sort of ironic popularity. You tracked him down, and when Pink saw him for the first time in over forty years, she ran and embraced him. He hugged her back. It took them a few hours of fawning over each other before she realized that nothing had happened when she touched him a second time. Something something true love something the power was within you the whole time?

But you had bigger fish to fry. The stewards of Orange's empire weren't too happy about their figurehead monarch suddenly rising from the dead, and for a while his position was precarious. He asked you to be his advisor, and you accepted. With your help, he was able to retake his throne. His first act was to fund research into the immortality serum you had heard about, which was discovered right on schedule in 2060.

The years went by. Orange's empire started colonizing new worlds, then new galaxies, until thousands of years later it changed its name to the Virgo Superconfederation. New people were born. New technologies were invented. New frontiers were conquered. Until finally, the stars started going out one by one.

Faced with the impending heat death, Orange elected to concentrate all his remaining resources here, on a single station in the center of the galaxy, which would wait out the final doom as long as possible. For billions of years, it burned through its fuel stockpile, until the final doom crept closer and closer.

And then a miracle occurred.

EPILOGUE

Red

This space station is AWESOME! There are lasers and holodecks and lots of HOT PUSSY! And all you have to do is turn a giant turbine for a couple of hours a day.

One of the eggheads in white coats tried to explain it to you once. He said that your BRUTE STRENGTH was some kind of scientific impossibility, because you didn't eat or drink any more than anyone else, and you didn't breathe in any more oxygen than anyone else, and you were actually kind of small and scrawny, but you were still strong enough and fast enough to turn a giant turbine thousands of times per minute.

He rambled on and on about thermodynamics. Said that every other process in the universe used at most as much energy as you put into it, but that your strength seemed almost limitless regardless of how much energy you took in as food. That made you special, somehow. It made you a "novel power source" that could operate "independently of external negentropy". You weren't sure what any of that meant, and honestly the scientist seemed sort of like a BETA CUCKOLD ORBITER to you. But

whatever was going on, they'd promised you that if you turned this turbine every day, you could have all the HOT PUSSY you wanted and be SUPER ALPHA.

You'd even met the head honcho once, a guy named King William. He told you that some of the energy you produced was going to power the station, but that the rest was going into storage. That over billions and billions of years, they would accumulate more and more stored negentropy, until it was enough to restart the universe. That it would be a cycle – a newborn universe lasting a few billion years, collapsing into a dark period when new negentropy had to be accumulated, followed by another universe again.

It all sounded way above your head. But one thing stuck with you. As he was leaving, the King remarked that it was ironic that when the black hole harvesters and wormholes and tachyon capacitors had all failed, it was a random really strong guy who had saved them.

You had always known, deep down, that BRUTE STRENGTH was what was really important. And here, at the end of all things, it is deeply gratifying to finally be proven right.