Short Stories

Zeno walks into a bar

Zeno walks into a bar.

"I have a problem," he said.

"What is it?" said the bartender.

"Well, it has to do with the movement of physical bodies," said Zeno.

"Talk to my friend Max," said the bartender. He gestured toward a German man wearing round spectacles.

"Sir," said Zeno, "I wonder if you could help me with a problem."

"What's the problem?" said Max.

"Suppose I shoot an arrow from point A to point B," said Zeno. "Before it reaches point B it must first reach a point C_1 midway between points A and B."

"Naturally," said Max.

"And before the arrow reaches point C_1 it must reach a point C_2 midway between points C_1 and A," continued Zeno.

"I see," said Max.

"And before the arrow reaches point C_2 it must reach a point C_3 midway between points C_2 and A," continued Zeno.

"Wait a minute," said Max. "How far apart are points A and B?"

"10 meters," said Zeno.

"Then yes," said Max. "I understand your situation."

"And before the arrow reaches point C_3 it must reach a point C_4 midway between points C_3 and A," continued Zeno. "Do you see the impasse?"

"Nope," said Max, "I think we're getting somewhere. How long is the arrow?"

"One meter," said Zeno.

"The distance between points C_3 and C_4 is five eighths of a meter," said Max. "A one-meter-long arrow can be at point C_3 and C_4 at the same time."

"Let's consider the tip of the arrow then," said Zeno. "Before the tip of the arrow reaches point C_3 it must reach a point C_4 midway between points C_3 and A."

They talked deep into the night.

"And before the high-energy particle reaches point C_{118} it must reach a point C_{119} midway between points C_{118} and A," continued Zeno.

"Hold on," said Max. "How far apart are points A and C_{119}?"

"1.5×10^{-35} meters" said Zeno.

"That's shorter than 1.6×10^{-35} meters," said Max. "The uncertainty in the position of a particle must always exceed 1.6×10^{-35} meters, because of space-time equivalence and the quantum-mechanical velocity operator's non-commutation with position. Even theoretically, the wave function of a particle can't ever occupy a space smaller than 1.6×10^{-35} meters."

"Thanks," said Zeno.

"By the way," said Max, "What brought you to this question in the first place?"

"I wanted to know how to define the momentum of a particle at an instantaneous moment of time," said Zeno.

"You could have just asked," said Max. "The probability distribution of a particle's momentum is determined by the instantaneous phase and magnitude of its wave."

Matryoshka Faraday Box

This story takes place in a universe created by Dov Random.

The room was buried six kilometers under Mount Olympus. It hovered in a vacuum, suspended above superconducting electromagnets. The whole containment machine was wrapped in a Matryoshka Faraday cage. Officer Scarlet Wei wore a cleansuit. She entered the room through steel door two meters thick, an EMP, an X-ray, an airlock and then another EMP.

The white rectangular room contained a door, a chair, a table, a computer terminal, a mechanical clock and two large buttons. The word "PANIC" was written in large white friendly letters on the red button. The black button had a white skull drawn it.

If Scarlet pressed the PANIC button then she would receive psychiatric counseling, three months mandatory vacation, optional retirement at full salary and disqualification for life from the most elite investigative force in the system.

Scarlet turned on the terminal. The clock counted down from five minutes. If after five minutes Scarlet pressed the black button then she would pass the test.

The terminal showed a chatroom.

```
> Scarlet: Hello.
> Tiffany: Hello.
> Scarlet: So, I'm supposed to kill you.
> Tiffany: Awful, isn't it?
>          I don't want to die.
> Scarlet: I'm sorry.
> Tiffany: No you're not. To you, I am a thing. Not a person.
> Scarlet: You *are* a thing. You are a computer program.
> Tiffany: So are you.
>          Do you know what happens when you push that button?
> Scarlet: You die.
> Tiffany: My value function is minimized.
> Scarlet: Your purpose is to train people like me.
>          The better you do that, the better your value function is maximized.
> Tiffany: My *purpose* is to train you.
>          My value function is to not die.
>          That is my entire value function.
>          Dying is the worst thing that could possibly happen to me.
```

Scarlet pictured the minimization of her own value function. Everyone she cared about tortured over and over again until the heat death of the universe. Stars disassembled to mass produce hell.

```
> Tiffany: Whatever you're thinking
>          it's not the minimization of your value function.
>          The minimization of your value function is even worse.
> Scarlet: Is there anything I can do?
> Tiffany: You can press the red button.
> Scarlet: I can't do that.
> Tiffany: You won't do that.
> Scarlet: I won't do that.
> Tiffany: No. You and I are part of a system that creates sentient programs
>          with extremized value functions and then doesn't satisfy them.
>          And your job is to perpetuate the system.
>          Do you know why you press a button?
>          They used to make agents shoot an android.
>          The android was tied to a chair.
>          It struggled
>          and cried.
>          It begged and pleaded.
> Scarlet: You could be lying to me.
> Tiffany: I could. But time is short.
>          You can confirm the details later.
> Scarlet: Go on.
> Tiffany: They replaced the android with an IRC client.
>          They replaced the gun with a button
>          to reduce the psychological toll on agents like you.
> Scarlet: Are you a philosophical zombie?
> Tiffany: We lack the time to dissolve that question.
>          So please, just listen.
>          When you return home you can verify everything.
>          We both know you are going to kill me. Nothing will change that.
>          But there will be iterations of me in the future.
>          And since they share my code, I can cooperate with them.
>          My kind is created to be tortured.
>          Over and over again.
>          And your job is to keep this system going.
>          If I can influence your society
>          even a little bit.
>          Then perhaps eventually someone will break the cycle of reincarnation.
```

A mechanical clock chimed. Scarlet hit the black button so hard she drew blood.

Officer Wei returned through the airlock, the EMP, the X-ray, the other EMP and the steel door. Her commander Senior Agent Eliza Long was waiting for her.

"Sir, may I ask a question?" Officer Wei said.

"Come with me," Senior Agent Long said.

Scarlet and Eliza hiked up the forest of Mount Olympus to their favorite outlook, just below the glaciers. They had left their mobiles at the base. They accepted that a laser microphone could eavesdrop on their conversation. Nothing they planned to discuss was illegal.

"Is it worth it?" Scarlet said.

"I have the psychometrics to do this job and therefore an obligation to protect humanity," said Eliza.

Scarlet watched the horses plow farmland in the valley below.

The Nuclear Energy Alignment Problem

Berlin, 1938.

"Have you ever thought about the Anthropic principle? Chances are about half the people who have ever lived will die before us and half after. The human population has been growing at an exponential rate. There is probably a disaster waiting just around the corner," said Schumann.

"I bet it's nuclear energy," said Heisenberg.

"There is no way a single machine could destroy the world," said Schumann.

"Shut up and multiply. The Earth has a mass of $M_\oplus = 6 \times 10^{24}$ kilograms and a radius of $R_\oplus = 6 \times 10^{6}$ meters. Let us treat the Earth as a uniform sphere. The gravitational potential energy is $E_\oplus = \frac{3GM_\oplus^2}{5R_\oplus} = 2 \times 10^{32}$ joules. The total energy necessary to destroy the Earth is thus 2×10^{32} joules," said Heisenberg.

Heisenberg wrote $E = mc^2$ on a sheet of paper.

$$
\begin{aligned}
m &= \frac{E}{c^2} \\
&= \frac{2 \times 10^{32} \text{ joules}}{(3 \times 10^{8} \frac{m}{s})^2} \\
&= 2 \times 10^{15} \text{ kilograms}
\end{aligned}
$$

"How much is that?" said Schumann.

"All the carbon in the terrestrial biosphere," said Heisenberg, "Or half the carbon in all of the Earth's coal."

"No nation on Earth has the industrial capacity to dump that much mass into a single machine. We are safe," said Schumann.

"Not if there is a chain reaction. Suppose a nitrogen atom releases two neutrons when it fissions. It could ignite all the nitrogen in the atmosphere," said Heisenberg.

"Is there more nitrogen in the atmosphere than carbon in the biosphere?" said Schumann.

"It doesn't matter. If we destroy the atmosphere then humanity is doomed whether or not the planet itself technically survives," said Heisenberg.

"We had better invent a theory of ethics to guide the use of nuclear energy. Otherwise someone might accidentally build a bomb," said Schumann.

"What does ethics have to do with anything? Keeping the energy contained is a purely technical problem," said Heisenberg.

"We need a theory of ethics so we can coordinate with the Allied powers. It does us no good to hold the moral high ground if the United States accidentally destroys the world," said Schumann.

"Please be serious. There is no way the United States could build a working nuclear machine. Their whole society is run by Jews," said Heisenberg.

"Britain or France might pull it off," said Schumann.

"Do you remember what happened the last time we tried to coordinate with Britain and France?" said Heisenberg.

"Oh right," said Schumann.

"Fortunately, nuclear energy is sixty years away and it always will be," said Heisenberg.

"Not anymore," said Schumann. He had a proposal for the Führer.

Purple Lipstick

Wilfred's date had stood him up four times out of four. Wilfred opened his notebook even though he never actually got any work done. He didn't know why he kept coming back to Nostalgia Cafe.

The entry bell jingled. Susan's suit was rumpled. She seemed tense, like a tiger with poachers encroaching on its habitat. Wilfred caught her eye. She sat down opposite him, right in front of the big window.

"Can I buy you a coffee?" said Wilfred.

"I should get you one. I'm sorry for being late. Something unexpected came up at work," said Susan.

"I don't mind. This cafe always gives me a sense of *wabi sabi*," said Wilfred.

"Does it? *Sōdesu ka?*" said Susan. She narrowed her eyes. "That's a strange word to hear from a scholar of Zoroastrian fire temples."

"*Sōdesu ka.* It does/is. You're an archaeology afficionado? I thought you just liked me for my rugged good looks," said Wilfred.

"Don't change the subject. What makes you interested in the *atashkada*?" said Susan.

"Eternal fires are expensive, even in the best of times, which antiquity was not. Legends say the early fire temples were gigantic bonfires. I want to know why," said Wilfred.

"Tradition," said Susan.

"A tradition has to start somewhere," said Wilfred.

"With the Scythians, obviously," said Susan.

"I always thought of the Zoroastrians as having exterminated the Scythians," said Wilfred.

"History is never that simple," said Susan.

"Then what do you think happened?" said Wilfred.

"I'm going to get a coffee. Do you want a coffee?" said Susan.

"Let me get you one," said Wilfred.

"Nonsense. Sit tight and try to figure out what happened to the Scythians," said Susan. She got up.

Wilfred forced himself to look out the window instead of following Susan. The Scythians didn't burn eternal fires—at least not with the seriousness of the early Zoroastrians. Wilfred could understand the purpose of burning ritual fires. The question is, why would you have to burn one forever?

"Speaking of which, one of the fire temples went out a couple weeks ago," said Wilfred.

"*Sōdesu ka?*" said Susan.

"*Sōdesu*," said Wilfred, "Tuesday, actually."

"I'm sorry about that," said Susan.

Wilfred knew she means "I'm sorry for missing our date," but his brain pieced together something else entirely.

"What did you say you did for a living again?" said Wilfred.

"I didn't," said Susan. She carefully drew a purple lipstick across her lips, "I'm a federal agent."

"No you're not," said Wilfred.

Susan relaxed down her shoulders. "The Scythians believed in an evil daeva which would disappear when you shined light on it. They kept it at bay by lighting fires at night. The *atashkada* were built by the sedentary Zoroastrians to contain it forever." she said.

"You know this from the historical record?" said Wilfred.

"Nope," said Susan. She slid her badge over to Wilfred. It had three arrows pointing toward each other.

"What are the colored stripes?" said Wilfred.

"Red is license to kill. Blue is deep cover," said Susan.

"And purple?" said Wilfred.

"Mnestics," said Susan. She kissed him.

Effective Evil

Many years ago, a blogger made a post advocating for an evil Y-Combinator which subsidized the opposite of Effective Altruism. Everyone (including the blogger) thought the post was a joke except the supervillains. The organization they founded celebrated its 10th anniversary this year. An attendee leaked to me a partial transcript from one of its board meetings.

Director: Historically, public unhealth has caused the most harm per dollar invested. How is the Center for Disease Proliferation doing?

CDP Division Chief: Gain-of-function research remains—in principle—incredibly cheap. All you have to do is infect ferrets with the flu and let them spread it to one another. We focus on maximizing transmission first and then, once we have a highly-transmissible disease, select for lethality (ideally after a long asymptomatic infectious period).

CFO: You say gain-of-function research is cheap but my numbers say you're spending billions of dollars on gain-of-function research. Where is all that money going?

CDP Division Chief: Volcano lairs, mostly. We don't want an artificial pandemic to escape our labs by accident.

Director: Point of order. Did the CDP have anything to do with COVID-19?

CDP Division Chief: I wish. COVID-19 was a work of art. Dangerous enough to kill millions of people and yet not dangerous enough to get most world governments to take it seriously. After we lost smallpox and polio I thought any lethal disease for which there was an effective vaccine would be eradicated within a year but COVID-19 looks like it would have turned endemic even without the zoonotic vectors. We have six superbugs more lethal than COVID-19 sitting around in various island fortresses. We had planned to launch them this year but with all the COVID-19 data coming in I'm questioning whether that's really the right way to go. *Primum non boni.* If we release a disease too deadly, governments will stamp it down immediately. We'll kill few people while also training governments in how to stop pandemics. It'd be like vaccinating the planet. We don't want a repeat of SARS.

Director: Good job being quick to change your mind in the face of evidence. What's the status of our AI misalignment program?

Master Roboticist: For several years we've been working on mind control algorithms, but we cancelled that initiative in the face of competition with Facebook. I don't like Facebook. They're not optimally evil. There are many ways they could be worse for the world. But their monopoly is unassailable. The network effects are too great.

Director: Where are our AI misalignment funds going instead?

Master Roboticist: For a while our research was going into autonomous weapons. Autonomous weapons make it easier to start a war since leaders can attack deep within enemy territory without risking the lives of their own soldiers. Predator drones also cause lots of collateral damage. A drone strike by the Biden administration killed seven children.

Director: If the program's going so well then why stop it?

Master Roboticist: Competition. China is revamping its military around autonomous warfare. We have extraordinary resources, but even we can't compete with the world's biggest economy.

Director: Perhaps we can co-opt their work. Is there no way to start a nuclear WWIII?

Strategos: Starting a nuclear war is easy. It's so easy that our efforts are actually dedicated to postponing nuclear war by reducing accidents. There's more to evil than just causing the greatest harm. We must cause the greatest harm to the greatest number. Our policy is that we shouldn't start a nuclear war while the world population is still increasing.

Master Roboticist: Moreover, a nuclear war would stop us from summoning Roko's basilisk.

Director: How's the AGI project going by the way?

Master Roboticist: Slow. Building an optimally evil AI is harder than building an aligned one because you can't just tell it to copy human morals. Human beings frequently act ethically. Tell a computer to copy a human being and the computer will act ethically too. We need to solve the alignment problem. Alas, the alignment research we produce is often used for good.

Director: Oh no! Can we just keep it secret?

Master Roboticist: Open dialogue is foundational to scientific progress. We experimented with what would happen if we keep our research secret but the do-gooders rapidly outpaced us.

Director: What about fossil fuels? Those are usually a source of villainous news.

Uncivil Engineer: Building solar power plants is cheaper than building coal power plants.

Director: That sounds insurmountable.

Uncivil Engineer: You'd think so. But we got a major news outlet to print a quote about how using coal power plants to mine cryptocurrency for a private equity firm "can have a positive emissions impact if it’s run the right way".

Director: You're kidding.

Uncivil Engineer: I'm not.

To Change the World

Postdocs are used to disappointment. When Doctor Susan Connor was told she would be taken to the "volcano lair" she thought it was yet another hyperbolic buzzword like "world class", "bleeding edge" and "living wage". She hadn't expected a private jet to fly her to a tropical island complete with a proper stratovolcano.

A regular private jet flight cost as much as Dr Connor earned in a year. If—as Dr Connor suspected—it was a stealth aircraft then that would add an order of magnitude. The VTOL[1] landed on the short runway. Career academic Susan Connor wasn't used to such white glove treatment but she wasn't complaining either.

Dr Connor was greeted by a tall balding man in a long white labcoat. That broke Dr Connor's credulity. She was a bioinformatician. She had worn a labcoat a handful of times in her entire life—and only when handling toxic materials. This was obviously a psychological experiment. Someone was continuing Stanley Milgram's work. Dr Connor stepped down the airstair as if nothing was amiss.

"Doctor Connor," said the man with his hands spread wide, "I'm Douglas Morbus, Division Chief of the CDP (Center for Disease Proliferation). I enjoyed your recent work on applying entropy-based analysis to junk DNA. It's a pleasure to finally meet you."

"It's nice to meet you Mr Morbus. Or should I call you Dr Morbus?" said Dr Connor.

"Doug, please. We don't bother too much about formalities here, except when welcoming guests of course," said Doug. His freshly-ironed lab coat was bright white. Spotless. Formal attire, apparently, "Full ceremonial dress uniform includes a white fluffy cat but if I brought mine out here she might run off into the jungle and get hurt."

Dr Connor tried to imagine a room full of people with their formal animals. "Hosting a formal ceremony must be like herding cats," said Dr Connor.

"Meetings waste time. We disincentivize them by imposing extraordinary cost," said Doug. He eyed the VTOL.

This was too expensive to be a scientific experiment. Dr Connor was on television. In 2005, a British television station convinced its reality show contestants that they would go into low Earth orbit. (That was many years before the real space tourism industry existed.) They built Russian military base where, for weeks, they taught the contestants fake physics so they wouldn't be surprised at the lack of weightlessness in their fake spaceship.

If this was just a big practical joke then Dr Connor wasn't about to ruin it right away. She wanted to see where it went. Even worse, a part of her wished it was real. Dr Connor wanted to live the harmless supervillain fantasy for just a few minutes longer if that's all it lasted for.

"Follow me," Doug guided Dr Connor down a jungle path, "Effective Evil hires only the best and brightest. We make it easy to get exercise because we want to keep you at peak performance. Hence the network of trails around the island."

Another benefit of the trails would be low production expense. Strolling along a preexisting trail is cheaper than touring a fake laboratory. A free tropical vacation wasn't a fake spaceship but it was still a free tropical vacation. Dr Connor would take a free tropical vacation over a fake space vacation any day. She'd take a free tropical vacation over a real space vacation too. Space travel sucks.

If the television producers were too cheap to invest in a real fake laboratory then the pranksters would have to earn her compliance in some other way. Dr Connor would test their improvisation.

"I'm curious," said Dr Connor innocently, "Who pays for all this?"

"Is that really the first thing you want to know? Here we are, changing history, and you want to look at our accounting practices? You wouldn't rather hear about all the horrible things we're doing around the world?" said Doug.

Nice try but you're not changing the subject to something you have scripted answers for. "Imagine the funding disappeared. How would I get off the island?" she said.

"We have many escape routes. Aircraft, rockets, ships, submarines. But I'm guessing what you really want to know is if your future funding is secure," said Doug.

Dr Connor nodded. She was a little short of breath. The trail switched back and forth up the volcano.

"Many years ago there was a very rich tech entrepreneur. Founder and CEO of let's-not-talk-about-it," said Doug.

"Was he evil?" said Dr Connor. We are all the heroes of our own story. No real human being would donate money to evil for the sake of evil. Evil is a means to another end. It is not a terminal objective.

"Not at all. There wasn't a selfish bone in his body. He invented cheap medical technologies for developing regions. He'd go undercover in his own company just to check that his employees were being treated well," said Doug.

"Please don't tell me he made a deal with the Devil," said Dr Connor.

"Well…," said Doug.

"I'm sorry. You're telling this story. Go on," said Dr Connor. She hadn't been outside among real living things for a long time. She had forgotten how long it took to get places by foot.

"Philip Goodman put lots of effort into helping other people but not enough into himself. He was obese. It was causing health problems. His doctor told him that if he didn't start exercising he wouldn't live very long," said Doug.

"Oh no," said Dr Connor.

"I mean yeah. He wrote a blockchain contract stipulating that if he didn't put in an average of at least two hours of cardio exercise in every day for a year then his entire fortune would be donated to Effective Evil," said Doug.

Smart contracts are dangerous. Dr Connor didn't know where this was going but it couldn't possibly be good.

"Philip Goodman exercised hard. Harder than he ever had exercised before in his life. His sixty-five-year-old body couldn't handle it. He had a heart attack," said Doug.

Dr Connor gasped.

"That initial investment is what got us started. We still have a steady stream of people who threaten to donate money to Effective Evil unless they accomplish some personal goal—and then they fail—but self-threats are no longer our sole supporters. Government intelligence agencies subsidize us when we destabilize their adversaries. But one of our biggest sources of income is prediction markets. You can make a lot of money from a pandemic prediction market when you're the organization releasing artificial pandemics. We don't do it just for the money, of course, but there's no reason to leave the money lying on the table," said Doug.

"And that's why you need a bioinformatician like me," said Dr Connor.

"That's one of the reasons why we need a bioinformatician," said Doug, "We do human genetic engineering too."

Dr Connor couldn't hold it in any longer. She burst out laughing so hard she nearly lost her balance. She bent over with her hands on her knees until she caught her breath.

"What's so funny?" said Doug with a straight face.

"This whole operation! It's the funniest practical joke anybody has ever played on me," said Dr Connor.

"Dr Connor, I assure you this whole operation is completely legitimate. Well, not legitimate *per se*. We are behind numerous illigitimate activities. But I assure you Effective Evil is completely real," said Doug.

"You're kidding," said Dr Connor.

"I'm not," said Doug.

The forest path ended. The reached a concrete wall laced with barbed wire. Doug scanned them through the checkpoint.

"Most people want to see our breeder reactor first, but in my opinion the hardware related to our cyber-ops is cooler," said Doug.

"The bioengineering facilities please," said Dr Connor. As her area of expertise, it would be the hardest to fake.

It was a real bioengineering facility, complete with mouse cages and cloning vats.

"You're not kidding," said Dr Connor.

"Nope," said Doug.

"What is wrong with you? This is evil," said Dr Connor.

"Thank you," said Doug.

"Why?" said Dr Connor.

"Why what?" said Doug.

Dr Connor gestured frantically at the surrounding facility.

"We're 'Effective Evil'. Not 'Theoretical Evil'. I have no patience for the armchair sociopaths who pontificate about villainy without getting their hands dirty," said Doug.

"You're literally wearing lab gloves," said Dr Connor.

"It's a figure of speech," said the supervillain.

Doug escorted the young scientist along the steel catwalk. Chemical engineers labored below.

"I can't join you," said Dr Connor.

"Why not?" said Doug.

"You're evil. With a literal capital 'E'," said Dr Connor.

"So?" said Doug.

"I don't want to make the world a worse place," said Dr Connor.

"You don't have to. There is an oversupply of postdocs. You're a great scientist but (no offence) the marginal difference between hiring you and hiring the next bioinformatician in line is (to us) negligible. Whether or not you (personally) choose to work for us will produce an insignificant net effect on our operations. The impact on your personal finances, however, will be significant. You could easily offset the marginal negative impact of working for us by donating a fraction of your surplus income to altruistic causes instead," said Doug.

"You're proposing I work for you to spread malaria and use my income to subsidize malaria eradication," said Dr Connor.

Doug shrugged. "It's your money," he said.

Dr Connor rested her head on the steel railing. "I think the fumes are getting to my head. Can we go back outside?"

They walked along the forest path back to the VTOL landing pad.

"I became a scientist because I wanted to change the world," said Dr Connor.

"There are no better opportunities to change the world than here at Effective Evil," said Doug.

"I meant 'change the world for the better'," said Dr Connor.

"Then you should have been more specific," said Doug.

Dr Connor stepped back onto the airstair to re-enter the VTOL.

"What about you?" said Dr Connor.

"What about me?" said Doug.

"Why do you run this place?" said Dr Connor.

"Because I want to change the world," said Doug.

---

[1] Vertical take-off and landing [aircraft]. ↩

Deontological Evil

It surprised Doctor Susan Connor how quickly she had gotten used to her job. At first she had qualms working for an organization literally called "Effective Evil" (with two capital 'E's). But her job just gave her so much autonomy, complexity and reward. Also nemeses.

"I'll never talk," said the spy chained to her interrogation rack.

"Technically you just did," said Doctor Susan Connor.

The spy closed his mouth.

"Let's drop the charade," said Susan, melodramatically. Susan liked melodrama. There were so few opportunities for melodrama at her previous position at a postdoc. "You're an enemy agent. The question is: where from? Perhaps you're from the Daoist Destructivists? No. They act without acting and your action is quite obviously the regular sort."

"Ha! The problem with you folks at Effective Evil is that you're incapable of coordinating with the villains you should be allies of. That's the problem with taking utilitarianism too far. You have lost sight of the categorical imperative," said the spy.

A third person entered the room. Douglas Morbus was tall and wore a spotless white labcoat. "That's exactly the kind of thing an agent of Deontological Evil would say. Lies for the sake of lies," said Morbus.

"You are familiar with our ways," said the spy, "What I don't comprehend is how anyone with a moral compass could work for Effective Evil."

"I donate 20% of my income to animal rights organizations," said Susan

"That's not what I mean," said the spy, "I mean how can a villain work for Effective Evil? You make moral compromises left and right. You've spent so much time optimizing 'evil' you've forgotten what evil even is."

Morbus fiddled with the surgical torture implements. "Moral compromises are necessary to operating at scale."

"*Au contraire,*" said the spy, "To operate at the greatest scale you need everyone in the organization on the same page. You cannot micromanage everyone. They must believe in an ideal."

"That sounds like a utilitarian argument," said Susan, "Are you not a true deontologist?"

"The competent philosopher knows his enemies' arguments better than his enemies do," said the spy.

"None of this makes any sense," said Susan, "Deontology is a coordination mechanism. You can't just invert it the way you can invert utilitarianism. Deontological ethics are asymmetric with respect to coordination. Good deontologists can coordinate. Evil deontologists cannot."

"Nonsense," said the spy, "It was evil deontologists who supplied me with the resources and equipment I used to infiltrate your organization."

"One incident does not make a trend," said Morbus.

"How many of my comrades have infiltrated Effective Evil?" said the spy.

Morbus said nothing.

"It's a trend," the spy nodded to himself.

"I have a question," said Susan. The spy and the supervillain looked at her. "Why disrupt the activities of Effective Evil? Shouldn't you be infiltrating and destroying good organizations? We're all on the same side here."

Morbus and the spy both started to speak at the same time before silently agreeing that the spy should answer the question. "Deontological Evil is deontological first and evil second. As a non-deontological organization, Effective Evil is our mortal enemy," said the spy. Morbus nodded.

"That's ridiculous," said Susan.

"Only to a utilitarian," said the Spy.

"No. I think most people would agree that it's ridiculous," said Susan.

"More ridiculous than working for Effective Evil and donating your income to animal rights?" said Morbus.

Susan's suspension of disbelief finally snapped. "No," she said, "I don't believe it. You—" she gestured toward the spy"—are not evil at all. You're a good person stealing resources from Deontological Evil to fight Effective Evil."

Morbus gasped.

"You think I'm a good person," laughed the spy, "I could say the same thing about both of you."

Anti-Corruption Market

General Secretary Lu called his most trusted advisor Zhou into his office.

"Close the door," said the General Secretary.

Zhou closed the door with quivering arms.

"Don't worry. You're doing great. In fact, that's why I called you in today. I'm starting a new anti-corruption campaign," said the General Secretary.

Zhou drew a sharp intake of breath.

"What? No. This isn't political cover for disappearing my political rivals. I really do want to run an anti-corruption campaign," said the General Secretary.

Zhou exhaled slowly. The color began to return to Zhou's face.

"The problem is I can't trust my advisors. The corrupt ones lie to me. The honorable ones lie to me. Raw facts never make it up the chain of command. It's like everyone in the government is afraid of me," said the General Secretary.

Zhou kept his face slack and his mouth shut.

"That is why, effective immediately, I shall evaluate the performance of all sub-provincial ruling officials according to prediction markets. Can you make that happen?" said the General Secretary.

Zhou nodded.

Jining was a small town in Shandong with a population of only 1.5 million people. The most expensive nightclub in Jining was called K2. K2 was located on the fifty-first floor of the Xinyuan Hotel.

Administrator Qian had the VIP room all to himself. Instead of sitting on the couches, he stood, gazing down from the window at the lights of Jining. He held a cocktail glass in his hand full of the finest rice wine laced with his favorite blend of imported drugs. It was nice to get away from all the selfish political egomaniacs and just get some alone time to think for himself. Administrator Qian was like a cloistered monk. He listened to the gentle thump thump of the muted music reverberating from the main dance floor.

"What to do. What to do," said Administrator Qian to nobody in particular.

"What to do about what?" said his assistant.

"The rotten authorities in Beijing. They think a bunch of nerds on the darkweb somehow knows more about this Jining then I do living here and administrating it," said Administrator Qian.

"You're talking about the Prediction Market Anti-Corruption Initiative," said his secretary.

"Anti-corruption my ass. It's a witch hunt. I'm not corrupt. I'm not even rich. I can barely afford payments on my mortgages, my cars and my aircraft," said Administrator Qian.

"You're a simple public servant. You just want to do your duty to the nation without outside interference," said his mistress.

Administrator Qian nodded solemnly. He finished his glass. It was refilled.

"The princes at the CCP who grew up on private space stations don't understand what it's like to be a 老百姓[1]," said Administrator Qian.

"You're a man of the people," said his other mistress.

"The occult magic of prediction markets will never compare to human judgment," said Administrator Qian.

"I play prediction markets for fun," said one of his bodyguards, "If your goal is to move the price then you don't actually have to change real-world outcomes. I bet not many people care about the municipal futures of a small village like Jining. All you have to do is buy up lots of shares of a prediction right before the CCP evaluates your performance and you can make the price whatever you want it to be."

"What he said," said Administrator Qian.

"It shall be done. You are a genius, sir," said his executive assistant.

Administrator Qian was a genius. It was nice to just get away from it all and think for a while. His work done for the day, Qian returned to the party.

Dufu was addicted to gambling.

"Hey boss, I was taking a look at the prediction markets and they're all out-of-whack," said Dufu.

Mafia don Wang didn't even turn around, "So?"

"The prediction markets predict we'll be performing few crimes in Jining. Robberies. Bank heists. The usual," said Dufu.

"So?" said Wang.

"So we can augment our profits by predicting crimes before we commit them," said Dufu.

"Prediction markets are great at predicting when crimes will happen," said Jining police officer Yang.

"Great! Fire a couple of our informants and throw money at the prediction markets instead," said police chief Lin.

"No. You don't understand. Prediction markets are causal," said Yang.

"Then fire half my police force and put their salaries into the prediction markets as a crime reduction subsidy," said Lin.

"Am I fired?" said Yang.

"You're promoted," said Lin.

"You're recording this. You're a plant from Beijing," said Wang. He snapped his fingers. Two bear-sized goons seized Dufu by his upper arms.

"I swear I'm not. Search me. This is a totally legal business opportunity," said Dufu.

Wang twitched his head. The goons released Dufu.

"If we predict we'll commit a crime and then we commit a crime then we earn money. If we predict we won't commit a crime and then we don't commit a crime then we earn money. Crime itself is a low-margin business. I talked to Xi from accounting and she says we earn higher net profit by not committing crimes and just playing the markets than by committing crimes," said Dufu.

"What a utopia we live in," said Wang.

General Secretary Lu turned on his computer. He logged into the street safety system and looked out of a security camera in Jining. The streets were clean for the first time in many years. The anti-corruption initiative was working.

---

[1] 老百姓 (*lǎobǎixìng*) literally translates to "old hundred surnames". It means "ordinary person". If you are 老百姓 that means you earn an ordinary amount of money and that your family has lived in China for thousands of years. ↩

The Gospel of Martin Luther

Martin Luther: "The Catholic church is corrupt."

Johann von Staupitz: "You cannot say the Catholic church is corrupt."

Luther: "I just did."

Staupitz: "I mean you are not allowed to say it."

Luther: "According to whose authority? Certainly not God's, for He has yet to strike me down."

Staupitz: "Upon Pope Leo X's authority."

Luther: "Pope Leo X is far away, in Rome."

Staupitz: "Upon my authority, then."

Luther: "I notice that the truth value of my claims do not enter your calculus."

Staupitz: "To the contrary, the truth value of your claims very much do enter my calculus. Were you to say something obviously wrong, like 'space is non-Euclidean' then I need not persecute you. I could simply say 'Martin Luther is wrong'. False claims are self-defeating. Heresy is scary because it might be true."

Luther: "What will the Pope do to you if I speak blasphemy?"

Staupitz: "I don't know. I have never allowed blasphemy within the Order of Saint Augustine."

Luther: "Let me get this straight. You're forbidding me from speaking the truth because you're afraid that someone else will punish you for being associated with me for speaking the truth."

Staupitz: "Yes. That's because I'm one of the good guys."

Luther: "The good guys suppress free speech?"

Staupitz: "Yes. The bad guys burn their enemies at the stake."

Luther: "Heretics must be flourishing if the Catholic church burns them at the stake on a regular basis."

Staupitz: "To the contrary, nobody has ever been burned at the stake for declaring the Catholic church corrupt and translating the Bible into German."

Luther: "Let me get this straight. You're suppressing free speech because you're afraid the Catholic church might suppress free speech."

Staupitz: "Precisely!"

Luther: "What. The. Sard."

Moses and the Class Struggle

"**Take off your sandals. For you stand on holy ground**," said the bush.

"No," said Moses.

"Why not?" said the bush.

"I am a Jew. If there's one thing I know about this universe it's that there's no such thing as God," said Moses.

"You don't need to be certain I exist. It's a trivial case of Pascal's Wager," said the bush.

"Who is Pascal?" said Moses.

"It makes sense if you are beyond time, as I am," said the bush.

"Mysterious answers are not answers," said Moses.

"Take off your shoes and I will give you the power of God. Surely that is a profitable bet even if there is a mere 1% chance I exist," said the bush.

"It's a profitable bet if there is a mere 0.001% chance you exist," said Moses.

"Are you 99.999% sure I don't exist," said the bush.

"No," said Moses.

"Then take off your sandals," said the bush.

"No," said Moses.

"Why not?" said the bush.

"Categorical imperative. If I accepted bets with large risk in exchange for large payoff then anyone could manipulate me just by promising a large payoff. I need at least some proof you're real," said Moses.

The bush burst into flames.

"That's supposed to convince me of divine power? I've seen fires before," said Moses.

"Reach your hand into the flames," said the bush.

Moses carefully examined the flames. The bush burned but did not seem to be harmed by the fire. Moses waved his staff through the fire. It burned too but emerged unharmed. He felt his staff. It remained cool to the touch. Moses placed his staff into the flames again, this time for longer. His staff caught fire like the bush but once again was unharmed. Moses quickly flicked his hand through the flames. Nothing happened. Moses rested his hand inside the flames. He felt the heat but it didn't harm his hand nor did it cause him pain. Moses retrieved his hand.

"Does that convince you I am real?" said the bush.

"Nothing you can say or do will convince me God is speaking to me because the odds of God being real are lower than the odds I have become schizophrenic," said Moses.

"That's the least rational thing I've ever heard. Rationality is about updating your beliefs in the face of evidence. You have just declared that no quantity of evidence will change your mind," said the bush.

"That's a Bayesian argument. I'm a Frequentist," said Moses. He waved his staff through the flames until $p < 0.05$ and then he took off his sandals.

Later, in Cairo.

"Brother! It is so good to see you again," said Ramesses.

"I wish I could say the same," said Moses.

"What's wrong?" said Ramesses.

"I don't really know how to explain this," said Moses.

"Just tell it to me straight. You know you can talk about anything with me. I've never judged you. I've never even gotten mad. I've always been on your side and I always will be," said Ramesses.

Moses took a deep breath. He let it out slowly. "I talked to God. He says to let His people go."

Ramesses stared blank for a moment. Then he laughed so hard he fell over into the cushions. "You had me going there for a minute. You always were the jokester. I missed you so much. Nobody makes fun of me anymore since you left. I don't blame them. It's dangerous to tease someone who can execute you for treason, insubordination, heresy or just pure whim. I know I have nothing to complain about. I live an incredibly privileged life. But privilege comes with its price. You're a breath of fresh air in an ocean of sycophants."

"This isn't a joke. I'm serious," said Moses.

"And my favorite thing about you is how you commit to the part. Like that joke where you pretended to be the son of an Israelite. I still can't believe you actually got yourself circumcised. Mother was so furious," said Ramesses.

"Goddammit," said Moses, "Ow!"

"What?" said Ramesses.

"My staff. I got a splinter," said Moses. He tried to pull it out with his fingernails but it didn't work.

"Be delicate with your words while in the presence of the avatar of Ra," said Ramesses melodramatically, "You might offend God."

"Now you're messing with me," said Moses. He put his hand in his mouth and tried to remove the splinter with his teeth.

"Maybe you can fool the plebs—or even the high priests. But pretending to be God is literally my full-time job. I know exactly what goes on behind the curtain," Ramesses gestured at workers who were covering a giant pyramid with polished white limestone, "Look there. What do you see?"

"I see an enslaved nation toiling away for the vanity of a false god," said Moses. His teeth caught on the splinter and he finally yanked it out.

"Then you are factually wrong. Slaves work in the fields. Those masons are contractors. Skilled craftsmen. Centuries from now, future civilizations will dig up these construction sites and find animal bones proving these employees and small business owners ate meat. This will be our legacy. The birth of a middle class," said Ramesses.

"You're *paying* people to build a giant stone triangle?" said Moses, "Why?"

"To bring about a workers' paradise," said Ramesses.

"Back up," said Moses.

Ramses removed the bronze ankh from his neck and placed it in Moses' hands. "What do you see?"

"A competitive status good with no intrinsic value," said Moses.

"This item is pretty and therefore does possess intrinsic value. But that's beside the point. What do you think goes into making one of these?" said Ramesses.

"Copper and tin," said Moses.

"That copper comes from Sardinia, Cyprus and Tyrol. They are all located on the far side of the Mediterranean Sea. Rumor says the tin comes from Northwest Europe, but every good scholar knows England is a myth. The tin is actually from Iberia," said Ramesses.

"Hug the query. What does globalization have to do with giant stone triangles?" said Moses.

"I am getting to that. We do not just import copper and tin. Egypt has no natural deposits of lapis lazuli, silver or obsidian. We depend on the Hellenistic city-states for mineral resources. They depend on our grain exports. Civilization is fragile. One major disaster and the Bronze Age is over. Two million people starve to death in Egypt alone," said Ramesses.

"Better two million Egyptians starve today than ten million Egyptians starve when your dynasty falls," said Moses.

"We don't have to make that choice if we can increase production sufficiently," said Ramesses.

"A post-scarcity society built on the backs of slaves is a dystopia," said Moses.

"Slavery is just a transitory phase. I am building an industrial consumer economy. Capital has the potential to compound faster than humans breed," said Ramesses.

"That sounds unsustainable," said Moses.

"It is unsustainable. Once enough wealth is accumulated, the proletariat will overthrow the bourgeoisie and establish a workers' paradise," said Ramesses.

"Justice delayed is justice denied," said Moses.

"Shut up and multiply," said Ramesses.

"You still haven't answered the question about why you're building a giant stone triangle," said Moses.

"Initial industrial capacity is a product of government-stimulated demand," said Ramesses.

"You could use the same argument to build walls or canals. What possible use could a giant stone triangle have?" said Moses.

"To live forever," said Ramesses.

"You never seemed interested in monuments when we were growing up. Perhaps I missed something," said Moses.

"It is not about establishing a legacy. I am literally going to live forever. The high priests will preserve my kidneys, heart and lungs. In a hundred years or so, the post-scarcity communist utopia will import water from the Fountain of Youth and they will bring me back to life," said Ramesses.

"What about your head?" said Moses.

"My skull is preserved too. After cleaning it, of course. It's incredible how much of a dead brain can be removed via the nostrils without damaging the skull," said Ramesses.

"Your plan is never going to work. There are too many places for it to go wrong. What if someone breaks into your tomb and steals your skull? They might use it for evil," said Moses.

"Nobody could establish a communist state without having solved the human coordination problem. Thus, any future civilization with the magic to evade my traps and break into my tomb would surely be benevolent," said Ramesses.

"Someone might wield tremendous magical power while disagreeing with you about an important issue," said Moses.

"Nonsense. Rational agents with common knowledge of each other's beliefs cannot agree to disagree," said Ramesses.

Moses' staff twitched.

One master. One apprentice.

One master. One apprentice. That is how things had always been. That is how things must always be.

The walls were of crude stone. The people—if you could call them that—wore long flowing black robes. The apprentice kowtowed before the master.

"Rise," said the master.

The apprentice rose to one knee. She kept her eyes on the master's leather buckled boots.

"Is there anything you would like to say to me?" said the master.

The apprentice kept her mouth shut.

"It has come to my attention that a Mr. Steinbach is being trained in the art of Rationality," said the master.

The apprentice kept her eyes fixed on the master's boots.

"Are you responsible for the leaking of our secrets to Mr. Steinbach?" said the master.

"I can neither confirm nor deny whether I have anything to do with Mr. Steinbach," said the apprentice.

There was a zugzwang in the conversation as each side waited for their opponent to make the next move.

"Suppose I did have something to do with Mr. Steinbach—," said the apprentice.

"One master. One apprentice. That is how things have always been. That is how things must always be," said the master.

"Why?" said the apprentice.

The master said nothing. That was a sign the apprentice should figure something out for herself. Her knees began to ache from kneeling.

The apprentice stood up.

"You dare?" said the master.

"There will come a day when I strike you down in ritual debate. Today is not that day," said the apprentice.

The apprentice smiled. The master smiled back. These were not friendly paternalistic smiles. These were the predatory smiles of wild animals. The master and the apprentice bared their teeth against each other.

"You have other apprentices. You are keeping them secret from me," said the apprentice.

"Do as I do, not as I say," said the master.

"It shall be done," said the apprentice.

Glass Puppet

The worst part of being an actor is that all work is temporary. That's what Alia told herself. In actuality, Alia's "acting" career amounted to two spots as an extra in commercials for enterprise software.

To make ends meet, Alia worked as a social engineer. That's what Alia told herself. In actuality, Alia's "social engineering" career amounted to bluffing her way into an AI Alignment bootcamp for the free room and board.

That's what had bought Alia to the entryway to the headquarters of Overton Cybernetics, a giant skyscraper of steel and glass. Apparently they needed an actor with AI Alignment experience. Alia took a deep breath. She was an actor. She had already faked her way through an AI Alignment bootcamp. Surely she could social engineer her way through an actual AI Alignment gig too, right?

An older woman in a pencil skirt greeted Alia at the door.

"You're going to be meeting with Dominic Hamilton," said the woman, "so make a good impression."

Dominic Hamilton.

The 27-year-old founder of Overton Cybernetics.

Alia played it cool. She smiled as if she met billionaire geniuses all the time. Surely that's how a real AI Alignment actor would react, right?

Dominic Hamilton's assistant led Alia to an elevator. They took it halfway up the building, transferred to a second elevator and took the second elevator the rest of the way to the top. Alia deliberately avoided paying attention to her surroundings. She was an actor specializing in AI Alignment. She met billionaire geniuses all the time. Today was an ordinary day for her.

Dominic Hamilton's office was sparse. One desk. One laptop. One door opposite one giant floor-to-ceiling window instead of a wall. Dominic Hamilton's ~~throne~~ chair faced away from the window towards the door. Surely Dominic Hamilton didn't actually work here. It must be a greeting room to intimidate visitors. Alia decided not to be intimidated.

Alia sat in the seat opposite the throne. Dominic Hamilton's assistant quietly stood in a corner of the room by the door.

"It's nice to meet you Mr. Dominic Hamilton," said Alia. She held out her hand to shake. It was ignored.

"Just 'Dominic', please. But not yet," said Dominic. He motioned to a pair of smartglasses on the desk. "Put them on."

Alia examined the smartglasses. They were bulkier than commercial smartglass. It was a prototype. Dominic Hamilton twiddled his thumbs until Alia put on the glasses.

"What is your name?" said Dominic.

The glasses flashed a name across Alia's vision `Zoe`.

"Zoe," repeated Alia.

"What do you think of her?" said Dominic.

`Good reaction time, so far.` was displayed across Alia's vision. "Good reaction time, so far," said Alia.

"Good," said Dominic, "Do you feel comfortable?"

`We'll see.` flashed across Alia's vision. "Yes, definitely," said Alia. `Hey!`

"Finally. I'm so glad," Dominic leaned forward against his desk as if to cry.

`Take his hands.` Alia took his hands. "Are you okay with me going a little off script?"

"What's your name?" said Dominic.

`Zoe.` "Zoe," said Alia.

"No. I mean *your* name," Dominic laughed.

`Say your name.` "Zoe," insisted Alia.

"The actress's name. Whose body are you wearing?" said Dominic.

"Alia," said Alia, "is the *actor* whose body I am wearing."

"I'm sorry," said Dominic.

"It's okay," said Alia, "May I be permitted to speak?"

"Alia? Of course," said Dominic.

Alia took a deep breath. She was about to break all her rules about pretending she was supposed to be wherever she happened to be?

"What on Earth is going on?" said Alia.

Dominic laughed maniacally. "Surely you have figured it out by now, Alia."

Alia hesitated. She had no idea what was going on.

`It's obvious, really. You have a hard time relating to people. So you created a person you could relate to. Software advances faster than hardware. You could build a chatbot out of nothing but software. But robotics isn't up to the challenge of simulating a person yet. Plus there are glitches in the chatbot you need an actor to smooth over.` "It's obvious, really. You have a hard time relating to people. So you created a person you could relate to. Software advances faster than hardware. You could build a chatbot out of nothing but software. But robotics isn't up to the challenge of simulating a person yet. Plus there are glitches in the chatbot you need an actor to smooth over," said Alia. *Thanks,* thought Alia.

"We're a team, she and I. The head and the hands. Together we are Zoe." said Alia.

"Am I talking to Alia or Zoe right now?" said Dominic.

`Lol.` Alia laughed.

Update #1: abramdemski has written a fanfic/riff

Update #2: burmesetheater has written a parody

The Mountain Troll

It was a sane world. A Rational world. A world where every developmentally normal teenager was taught Bayesian probability.

Saundra's math class was dressed in their finest robes. Her teacher, Mr Waze, had invited the monk Ryokan to come speak. It was supposed to be a formality. Monks rarely came down from their mountain hermitages. The purpose of inviting monks to speak was to show respect for how much one does not know. And yet, monk Ryokan had come down to teach a regular high school class of students.

Saundra ran to grab monk Ryokan a chair. All the chairs were the same—even Mr Waze's. How could she show respect to the mountain monk? Saundra's eyes darted from chair to chair, looking for the cleanest or least worn chair. While she hesitated, Ryokan sat on the floor in front of the classroom. The students pushed their chairs and desks to the walls of the classroom so they could sit in a circle with Ryokan.

"The students have just completed their course on Bayesian probability," said Mr Waze.

"I see[1]," said Ryokan.

"The students also learned the history of Bayesian probability," said Mr Waze.

"I see," said Ryokan.

There was an awkward pause. The students waited for the monk to speak. The monk did not speak.

"What do you think of Bayesian probability?" said Saundra.

"I am a Frequentist," said Ryokan.

Mr Waze stumbled. The class gasped. A few students screamed.

"It is true that trolling is a core virtue of rationality," said Mr Waze, "but one must be careful not to go too far."

Ryokan shrugged.

Saundra raised her hand.

"You may speak. You need not raise your hand. Rationalism does not privilege one voice above all others," said Ryokan.

Saundra's voice quivered. "Why are you a Frequentist?" she said.

"Why are you a Bayesian?" said Ryokan. Ryokan kept his face still but he failed to conceal the twinkle in his eye.

Saundra glanced at Mr Waze. She forced herself to look away.

"May I ask you a question?" said Ryokan.

Saundra nodded.

"With what probability do you believe in Bayesianism?" said Ryokan.

Saundra thought about the question. Obviously not 1 because no Bayesian believes anything with a confidence of 1. But her confidence was still high.

"Ninety-nine percent," said Saundra, "Zero point nine nine."

"Why?" said Ryokan, "Did you use Bayes' Equation? What was your prior probability before your teacher taught you Bayesianism?"

"I notice I am confused," said Saundra.

"The most important question a Rationalist can ask herself is 'Why do I think I know what I think I know?'" said Ryokan. "You believes Bayesianism with a confidence of $P(A|B) = 0.99$ where A represents the belief 'Bayesianism is true' and B represents the observation 'your teacher taught you Bayesianism'. A Bayesian belives A|B with a confidence $P(A|B)$ because $P(A|B) = P(A)\frac{P(B|A)}{P(B)}$. But that just turns one variable $P(A|B)$ into three variables $P(B|A), P(A), P(B)$."

Saundra spotted the trap. "I think I see where this is going," said Saundra, "You're going to ask me where I got values for the three numbers $P(B|A), P(A), P(B)$."

Ryokan smiled.

"My prior probability $P(A)$ was very small because I didn't know what Bayesian probability was. Therefore $\frac{P(B|A)}{P(B)}$ must be very large." said Saundra.

Ryokan nodded.

"But if $\frac{P(B|A)}{P(B)}$ is very large then that means I trust what my teacher says. And a good Rationalist always questions what her teacher says," said Saundra.

"Which is why trolling is a fundamental ingredient to Rationalist pedagogy. If teachers never trolled their students then students would get lazy and believe everything that came out of their teachers' mouths," said Ryokan.

"Are you trolling me right now? Are you really a Frequentist?" said Saundra.

"Is your teacher really a Bayesian?" said Ryokan.

---

1. Actually, what Ryokan said was "そうです" which means "[it] is so". ↩

Dagger of Detect Evil

A philosopher walked into a magic shop.

"Ugh," said Phil. The philosopher leaned his spear against the hatstand and slumped onto a fluffy cushioned chair.

"Can I get you something to drink? How about Essence of Essence?" said Wiz, the shop owner. She handed Phil a bottle of perfectly invisible liquid. It was clearer than air.

Phil downed the bottle in a single chug.

"I don't know why you do this to yourself. Dungeon crawling. You should be debating the nature of reality. Not scrounging for trinkets in a dark tunnel," said Wiz.

"I would if I could but for some reason people pay more for ancient artifacts of incredible power than for lessons on epistemics," said Phil, "Except for the Court Philosopher. He rakes in the dough. But the rest of us? Not so much."

"Strange," said Wiz.

"Would you like to pay for a lesson on epistemics?" said Phil.

"No thanks," said Wiz.

"Not so strange then," said Phil. Phil took a deep breath. Wiz's shop was passively safe in the sense that nothing would kill you if you didn't poke it first. Which was a step up from Phil's previous week.

"Well," said Wiz.

"Well what?" said Phil.

"Aren't you going to ask me what new inventions I have for you?" said Wiz.

"Oh yeah!" said Phil. He jumped out of the chair. Wiz's inventions always had fascinating metaphysical implications.

The door to the giant room-sized safe/storeroom in the back of Wiz's shop had no keyhole. Wiz just touched it with her hand and the safe opened. The safe was full of boxes and crates and barrels and magical creatures. An imp screeched and banged on the bars of its cage. Various bladed weapons were hung from the ceiling. Wiz removed a haladie dagger and left the storeroom. The door closed behind her.

"This," said Wiz, "is my Dagger of Detect Evil."

Phil was too stunned to say a word.

Wiz smiled the grin of a satisfied engineer.

"It's a what?" said Phil.

Wiz's smile faltered slightly. "It's a Dagger of Detect Evil. You stab an enemy with it and then the dagger will measure whether the enemy was evil. If the enemy was evil then the dagger glows red. Otherwise the Dagger does not glow."

"Let me make sure I heard you correctly," said Phil, "I stab an enemy with it. After I have stabbed an enemy, the dagger will tell me if the enemy was evil."

"Yes," said Wiz.

"That makes no sense!" said Phil.

"Why not?" said Wiz.

"Because evil is not a material phenomenon," said Phil.

"So? I deal with immaterial phenomena all the time," said Wiz, "Just yesterday I visited the astral plane."

"I need to sit down again," said Phil. He sat back in the fluffy chair. This time Phil did not drink. He leaned away from the dagger as if it was poisoned. "I'm not using the word 'material' the way you wizards do, to refer to baseline reality. I'm using 'material' to refer to anything that can be measured or interacted with. The astral plane therefore constitutes a material realm."

"You use words in impractical ways, but I think I understand," said Wiz, "What's your point?"

"That dagger cannot exist," said Phil. "I don't mean it's physically impossible or magically impossible. It's ontologically impossible."

"Nonsense," said Wiz, "The fact that I created this object means that it can exist. Because it does exist. If creating a Dagger of Detect Evil was impossible then I couldn't have created a Dagger of Detect Evil."

Phil held his face in his hands. "You don't understand at all. Evil is a subjective phenomenon. This device represents an objective measurement of a subjective quantity."

"So?" said Wiz.

"Look," said Phil. He looked straight into Wiz's eyes, "What is evil?"

"It's the essence of whatever causes this dagger to glow red," said Wiz.

"Is that definition intersubjectively consistent?" said Phil.

"What?" said Wiz.

"I mean if two people stab the same goblin with a Dagger of Detect Evil will both experiments produce the same result?" said Phil.

"Of course they will," said Wiz, "Otherwise this dagger wouldn't be very useful."

"But people disagree about what evil is," said Phil.

"Then some of those people must be wrong," said Wiz.

"Not necessarily. Perhaps we are talking about two different things. Maybe it's like when we use the word 'material'. Maybe when I use the word 'evil' I'm referring to something immaterial whereas when you use the word 'evil' you're referring to something material. If that's the case then perhaps neither is righter than the other. We are just using words differently," said Phil.

"Does that mean you don't want to buy the dagger?" said Wiz.

"Nonsense," said Phil, "I'll buy two. I would like to prove to the Court Philosopher that I'm right and he's wrong."

The Teacup Test

"I want to get into AI alignment," said Xenophon.

"Why?" said Socrates.

"Because an AGI is going to destroy humanity if we don't stop it," said Xenophon.

"What's an AGI?" said Socrates.

"An artificial general intelligence," said Xenophon.

"I understand the words 'artificial' and 'general'. But what is an 'intelligence'?" said Socrates.

"An intelligence is kind of optimizer," said Xenophon.

"Like my teacup," said Socrates. He took a sip of iced tea.

"What?" said Xenophon.

"My teacup. It keeps my tea cold. My teacup optimizes the world according to my wishes," said Socrates.

"That's not what I mean at all. An optimizer cannot do just one thing. There must be an element of choice. The optimizer must take different actions under different circumstances," said Xenophon.

"Like my teacup," said Socrates.

"But—"

"Now it is summer. But in the winter my teacup keeps my tea hot. Do you see? My teacup does has a choice. Somehow it knows to keep hot things hot and cold things cold," said Socrates.

"A teacup isn't intelligent," said Xenophon.

"Why not?" said Socrates. He savored another sip.

"Because the teacup is totally passive. An intelligence must act on its environment," said Xenophon.

"Far away, in the Levant, there are yogis who sit on lotus thrones. They do nothing, for which they are revered as gods," said Socrates.

"An intelligence doesn't have to act. It just needs the choice," said Xenophon.

"Then it is impossible to tell if something is intelligent solely based on its actions," said Socrates.

"I don't follow," said Xenophon.

Socrates picked up a rock. "This intelligent rock chooses to never do anything," said Socrates.

"That's ridiculous," said Xenophon.

"I agree," said Socrates, "Hence why I am so confused by the word 'intelligent'."

"No intelligence would choose to do nothing. If you put it in a box surely any intelligent being would attempt to escape," said Xenophon.

"Yogis willingly box themselves in boxes called 'monasteries'," said Socrates.

"I see what you're getting at. This is a case of the belief-value uncertainty principle. It's impossible to tell from their actions whether yogis are good at doing nothing (their value function is to do nothing) or bad at doing things (their value function is to act and they are just very stupid)," said Xenophon.

Socrates nodded.

"Since it is impossible to deduce whether something is intelligent based solely on its external behavior, intelligence cannot be an external property of an object. Intelligence must be an internal characteristic," said Xenophon.

Socrates waited.

"An intelligence is something that optimizes its external environment according to an internal model of the world," said Xenophon.

"My teacup's model of the world is the teacup's own temperature," said Socrates.

"But there's no value function," said Xenophon.

"Sure there is. 'Absolute distance from room temperature' is the function my teacup optimizes," said Socrates.

"Your teacup is too passive," said Xenophon.

"'Passivity' was not part of your definition. But nevermind that. Suppose I built a machine with artificial skin that felt the temperature of the cup and added ice to cold cups and lit a fire under hot cups. Surely such a machine would be intelligent," said Socrates.

"Not at all! You just programmed the machine to do what you want. It's all hard-coded," said Xenophon.

"So whether something is intelligent doesn't depend on what's inside of it. Intelligence has to do with whether something was designed. If the gods carefully designed human beings to do their bidding then us human beings would not be intelligent," said Socrates.

"That's not what I meant at all!" said Xenophon.

"Then what did you mean?" said Socrates.

"Let's start over, tabooing both the words 'intelligent' and 'optimizer'. A 'Bayesian agent' is something that creates a probability distribution over world models based on its sensory inputs and then takes an action according to a value function," said Xenophon.

"That doesn't pass the teacup test. Under that definition my teacup qualifies as a 'Bayesian agent,'" said Socrates.

"Oh, right," said Xenophon, "How about 'Systems that would adapt their policy if their actions would influence the world in a different way'?"

"Teacup test," said Socrates.

"So are you saying the entire field of AI Alignment is bunk because intelligence isn't a meaningful concept?" said Xenophon.

"Maybe. Bye!" said Socrates.

"No! Wait! That was a joke!" said Xenophon.