



# Positivism and Self Deception

1. [Simultaneously Right and Wrong](#)
2. [The Mystery of the Haunted Rationalist](#)
3. [The Apologist and the Revolutionary](#)
4. [You May Already Be A Sinner](#)
5. [Talking Snakes: A Cautionary Tale](#)
6. [The Skeptic's Trilemma](#)
7. [Are You a Solar Deity?](#)
8. [The "Spot the Fakes" Test](#)
9. [How to Not Lose an Argument](#)
10. [The Power of Positivist Thinking](#)
11. [When Truth Isn't Enough](#)

# Simultaneously Right and Wrong

**Related to:** [Belief in Belief](#), [Convenient Overconfidence](#)

*"You've no idea of what a poor opinion I have of myself, and how little I deserve it."*

-- W.S. Gilbert

In 1978, Steven Berglas and Edward Jones performed a [study](#) on voluntary use of performance inhibiting drugs. They asked subjects to solve certain problems. The control group received simple problems, the experimental group impossible problems. The researchers then told all subjects they'd solved the problems successfully, leaving the controls confident in their own abilities and the experimental group privately aware they'd just made a very lucky guess.

Then they offered the subjects a choice of two drugs to test. One drug supposedly enhanced performance, the other supposedly handicapped it.

There's a cut here in case you want to predict what happened.

---

Males in the lucky guesser group chose the performance-inhibiting drug significantly more than those in the control group<sup>1</sup>.

The researchers conjectured that the lucky guessers felt good about acing the first test. They anticipated failing the second whether they chose the helpful or the harmful drug. So they chose the harmful drug as an excuse: "Oh, I would have passed the test, only the drug was making me stupid." As the study points out, this is a win-win situation: if they fail, the drug excuses their failure, and if they succeed it's doubly impressive that they passed even with a handicap.

Since this study, psychologists have applied self-handicapping to interpret common phenomena like [procrastination](#), [hypochondria](#), and [minority underachievement](#); they've also discovered factors that increase or decrease self-handicapping tendencies (for example, you can increase it by making the test more relevant to your self-esteem: just say "this is a proven test of intellectual ability in general").

But some especially interesting studies investigated the effect of privacy on self-handicapping. For example, Hobden conducted an [experiment](#) similar to Berglas and Jones', albeit with tapes of performance-enhancing or -handicapping music instead of drugs. The twist: half the subjects' choice of tapes and test scores were public, and half the subjects believed their choice of tapes and test scores were anonymous. What happens when no one but the subject himself will ever know his test score? He self-handicaps just as often as everyone else. And it seems to *\*work\**. The same set of studies showed that subjects who self-handicap on a test are less likely to attribute their failure on the test to their own incompetence.

In order to handicap, subjects must have an inaccurate assessment of their own abilities. Otherwise, there's no self-esteem to protect. If I believe my IQ is 80, and I get 80 on an IQ test, I have no incentive to make excuses to myself, or to try to explain away the results. The only time I would want to explain away the results as based on some external factor was if I'd been going around thinking my real IQ was 100.

But subjects also must have an accurate assessment of their own abilities. Subjects who take an easy pre-test and expect an easy test do not self-handicap. Only subjects who understand their low chances of success can think "I will probably fail this test, so I will need an excuse"<sup>2</sup>.

If this sounds familiar, it's because it's another form of the dragon problem from [Belief in Belief](#). The believer says there is a dragon in his garage, but expects all attempts to detect the dragon's presence to fail. Eliezer writes: "The claimant must have an accurate model of the situation somewhere in his mind, because he can anticipate, in advance, exactly which experimental results he'll need to excuse."

Should we say that the subject believes he will get an 80, but believes in believing that he will get a 100? This doesn't quite capture the spirit of the situation. Classic belief in belief seems to involve value judgments and complex belief systems, but self-handicapping seems more like simple overconfidence bias<sup>3</sup>. Is there any other evidence that overconfidence has a belief-in-belief aspect to it?

Last November, [Robin described](#) a study where subjects were less overconfident if asked to predict their performance on tasks they will actually be expected to complete. He ended by noting that "It is almost as if we at some level realize that our overconfidence is unrealistic."

Belief in belief in religious faith and self-confidence seem to be two areas in which we can be simultaneously right and wrong: expressing a biased position on a superficial level while holding an accurate position on a deeper level. The specifics are different in each case, but perhaps the same general mechanism may underlie both. How many other biases use this same mechanism?

## Footnotes

1: In most studies on this effect, it's most commonly observed among males. The reasons are too complicated and controversial to be discussed in this post, but are left as an exercise for the reader with a background in evolutionary psychology.

2: Compare the ideal Bayesian, for whom expected future expectation is always the same as the current expectation, and investors in an ideal stock market, who must always expect a stock's price tomorrow to be on average the same as its price today - to this poor creature, who accurately predicts that he will lower his estimate of his intelligence after taking the test, but who doesn't use that prediction to change his pre-test estimates.

3: I have seen "overconfidence bias" used in two different ways: to mean poor calibration on guesses (ie predictions made with 99% certainty that are only right 70% of the time) and to mean the tendency to overestimate one's own good qualities and chance of success. I am using the latter definition here to remain consistent with the common usage on Overcoming Bias; other people may call this same error "optimism bias".

# The Mystery of the Haunted Rationalist

Followup to: [Simultaneously Right and Wrong](#)

*"The most merciful thing in the world, I think, is the inability of the human mind to correlate all its contents."*

- H.P. Lovecraft, [The Call of Cthulhu](#)

There is an old yarn about two skeptics who stayed overnight in a supposedly haunted mansion, just to prove they weren't superstitious. At first, they laughed and joked with each other in the well-lit master bedroom. But around eleven, there was a thunderstorm - hardly a rare event in those parts - and all the lights went off. As it got later and later, the skeptics grew more and more nervous, until finally around midnight, the stairs leading up to their room started to creak. The two of them shot out of there and didn't stop running until they were in their car and driving away.

So the skeptics' emotions overwhelmed their rationality. That happens all the time. Is there any reason to think this story proves anything more interesting than that some skeptics are cowards?

The Koreans have a superstition called "[fan death](#)": if you sleep in a closed room with a fan on all night, you will die. Something about the fan blades shredding the oxygen molecules or something. It all sounds pretty far-fetched, but in Korea it's endorsed by everyone from doctors to the government's official consumer safety board.

I don't believe in ghosts, and I don't believe in fan death. But my reactions to spending the night in a haunted mansion and spending the night with a fan are completely different. Put me in a haunted mansion, and I'll probably run out screaming the first time something goes bump in the night<sup>1</sup>. Put me in a closed room with a fan and I'll shrug and sleep like a baby. Not because my superior rationality has conquered my fear. Because fans just plain don't kill people by chopping up oxygen, and to think otherwise is just *stupid*.

So although it's correct to say that the skeptics' emotions overwhelmed their rationality, they wouldn't have those emotions unless they thought on *some* level that ghosts were worth getting scared about.

A psychologist armed with the theory of belief-profession versus [anticipation-control](#) would conclude that I profess disbelief in ghosts to fit in with my rationalist friends, but that I anticipate being killed by a ghost if I remain in the haunted mansion. He'd dismiss my skepticism about ghosts as exactly the same sort of belief in belief afflicting the man who thinks his dragon is [permeable to flour](#).

If this psychologist were really interested in investigating my beliefs, he might offer me X dollars to stay in the haunted mansion. This is all a thought experiment, so I can't say for certain what I would do. But when I imagine the scenario, I visualize myself still running away when X = 10, but fighting my fear and staying around when X = 1000000.

This looks suspiciously like I'm making an expected utility calculation. Probability of being killed by ghost \* value of my life, compared to a million dollars. It also looks like I'm using a rather high number for (probability of being killed by ghost): certainly still less than .5, but much greater than the  $<.001$  I would consciously assign it. Is my mind haunted by an invisible probability of ghosts, ready to jump out and terrify me into making irrational decisions?

How can I defend myself against the psychologist's accusation that I merely profess a disbelief in ghosts? Well, while I am running in terror out of the mansion, a bookie runs up beside me. He offers me a bet: he will go back in and check to see if there is a ghost. If there isn't, he owes me \$100. If there is, I owe him \$10,000 (payable to his next of kin). Do I take the bet?

Thought experiments don't always work, but I imagine myself taking the bet. I assign a less than 1/100 chance to the existence of ghosts, so it's probably a good deal. The fact that I am running away from a ghost as I do the calculation changes the odds not at all.

But if that's true, we're now up to three different levels of belief. The one I profess to my friends, the one that controls my anticipation, and the one that influences my emotions.

There are no ghosts, profess skepticism.

There are no ghosts, take the bet.

There are ghosts, run for your life!

### **Footnote**

1: I worry when writing this that I may be alone among Less Wrong community members, and that the rest of the community would remain in the mansion with minimal discomfort. If "run screaming out of the mansion" is too dramatic for you, will you agree that you might, after the floorboards get especially creaky, feel a tiny urge to light a candle or turn on a flashlight? Even that is enough to preserve the point I am trying to make here.



# The Apologist and the Revolutionary

Rationalists complain that most people are too willing to [make excuses](#) for their positions, and too unwilling to abandon those positions for ones that better fit the evidence. And most people really *are* pretty bad at this. But certain stroke victims called anosognosiacs are much, much worse.

[Anosognosia](#) is the condition of not being aware of your own disabilities. To be clear, we're not talking minor disabilities here, the sort that only show up during a comprehensive clinical exam. We're talking paralysis or even blindness<sup>1</sup>. Things that should be pretty hard to miss.

Take the example of the woman discussed in Lishman's [Organic Psychiatry](#). After a right-hemisphere stroke, she lost movement in her left arm but continuously denied it. When the doctor asked her to move her arm, and she observed it not moving, she claimed that it wasn't actually her arm, it was her daughter's. Why was her daughter's arm attached to her shoulder? The patient claimed her daughter had been there in the bed with her all week. Why was her wedding ring on her daughter's hand? The patient said her daughter had borrowed it. Where was the patient's arm? The patient "turned her head and searched in a bemused way over her left shoulder".

Why won't these patients admit they're paralyzed, and what are the implications for neurotypical humans? Dr. [Vilayanur Ramachandran](#), leading neuroscientist and current holder of the world land-speed record for hypothesis generation, has a theory.

One immediately plausible hypothesis: the patient is unable to cope psychologically with the possibility of being paralyzed, so he responds with denial. Plausible, but according to Dr. Ramachandran, wrong. He notes that patients with left-side strokes almost never suffer anosognosia, even though the left side controls the right half of the body in about the same way the right side controls the left half. There must be something special about the right hemisphere.

Another plausible hypothesis: the part of the brain responsible for thinking about the affected area was damaged in the stroke. Therefore, the patient has lost access to the area, so to speak. Dr. Ramachandran doesn't like this idea either. The lack of right-sided anosognosia in left-hemisphere stroke victims argues against it as well. But how can we disconfirm it?

Dr. Ramachandran performed an experiment<sup>2</sup> where he "paralyzed" an anosognosiac's good right arm. He placed it in a clever system of mirrors that caused a research assistant's arm to look as if it was attached to the patient's shoulder. Ramachandran told the patient to move his own right arm, and the false arm didn't move. What happened? The patient claimed he could see the arm moving - a classic anosognosiac response. This suggests that the anosognosia is not specifically a deficit of the brain's left-arm monitoring system, but rather some sort of failure of rationality.

Says Dr. Ramachandran:

The reason anosognosia is so puzzling is that we have come to regard the 'intellect' as primarily propositional in character and one ordinarily expects propositional logic to be internally consistent. To listen to a patient deny ownership of her arm and yet, in the same breath, admit that it is attached to her

shoulder is one of the most perplexing phenomena that one can encounter as a neurologist.

So what's Dr. Ramachandran's solution? He [posits two different reasoning modules](#) located in the two different hemispheres. The left brain tries to fit the data to the theory to preserve a coherent internal narrative and prevent a person from jumping back and forth between conclusions upon each new data point. It is primarily an apologist, there to explain why any experience is exactly what its own theory would have predicted. The right brain is the seat of the [second virtue](#). When it's had enough of the left-brain's confabulating, it initiates a Kuhnian paradigm shift to a completely new narrative. Ramachandran describes it as "a left-wing revolutionary".

Normally these two systems work in balance. But if a stroke takes the revolutionary offline, the brain loses its ability to change its mind about anything significant. If your left arm was working before your stroke, the little voice that ought to tell you it might be time to reject the "left arm works fine" theory goes silent. The only one left is the poor apologist, who must tirelessly invent stranger and stranger excuses for why all the facts really fit the "left arm works fine" theory perfectly well.

It gets weirder. For some reason, [squirting cold water into the left ear canal](#) wakes up the revolutionary. Maybe the intense sensory input from an unexpected source makes the right hemisphere unusually aroused. Maybe distorting the balance sense causes the eyes to move rapidly, activating a latent system for inter-hemisphere co-ordination usually restricted to REM sleep<sup>3</sup>. In any case, a patient who has been denying paralysis for weeks or months will, upon having cold water placed in the ear, admit to paralysis, admit to having been paralyzed the past few weeks or months, and express bewilderment at having ever denied such an obvious fact. And then the effect wears off, and the patient not only denies the paralysis but denies ever having admitted to it.

This divorce between the apologist and the revolutionary might also explain some of the odd behavior of [split-brain](#) patients. Consider [the following experiment](#): a split-brain patient was shown two images, one in each visual field. The left hemisphere received the image of a chicken claw, and the right hemisphere received the image of a snowed-in house. The patient was asked verbally to describe what he saw, activating the left (more verbal) hemisphere. The patient said he saw a chicken claw, as expected. Then the patient was asked to point with his left hand (controlled by the right hemisphere) to a picture related to the scene. Among the pictures available were a shovel and a chicken. He pointed to the shovel. So far, no crazier than what we've come to expect from neuroscience.

Now the doctor verbally asked the patient to describe why he just pointed to the shovel. The patient verbally (left hemisphere!) answered that he saw a chicken claw, and of course shovels are necessary to clean out chicken sheds, so he pointed to the shovel to indicate chickens. The apologist in the left-brain is helpless to do anything besides explain why the data fits its own theory, and its own theory is that whatever happened had something to do with chickens, dammit!

The logical follow-up experiment would be to ask the right hemisphere to explain the left hemisphere's actions. Unfortunately, the right hemisphere is either non-linguistic or as close as to make no difference. Whatever its thoughts, it's keeping them to itself.

...you know, my mouth is *still* agape at that whole cold-water-in-the-ear trick. I have this fantasy of gathering all the leading creationists together and squirting ice cold



water in each of their left ears. All of a sudden, one and all, they admit their mistakes, and express bafflement at ever having believed such nonsense. And then ten minutes later the effect wears off, and they're all back to talking about irreducible complexity or whatever. I don't mind. I've already run off to upload the video to YouTube.

This is surely so great an exaggeration of Dr. Ramachandran's theory as to be a parody of it. And in any case I don't know how much to believe all this about different reasoning modules, or how closely the intuitive understanding of it I take from his paper matches the way a neuroscientist would think of it. Are the apologist and the revolutionary active in normal thought? Do anosognosiacs demonstrate the same pathological inability to change their mind on issues other than their disabilities? What of the argument that [confabulation](#) is a rather common failure mode of the brain, shared by some conditions that have little to do with right-hemisphere failure? Why does the effect of the cold water wear off so quickly? I've yet to see any really satisfying answers to any of these questions.

But whether Ramachandran is right or wrong, I give him enormous credit for doing serious research into the neural correlates of human rationality. I can think of few other fields that offer so many potential benefits.

## Footnotes

1: See [Anton-Babinski syndrome](#)

2: See Ramachandran's "The Evolutionary Biology of Self-Deception", the link from "posits two different reasoning modules" in this article.

3: For Ramachandran's thoughts on REM, again see "The Evolutionary Biology of Self Deception"

# You May Already Be A Sinner

**Followup to:** [Simultaneously Right and Wrong](#)

**Related to:** [Augustine's Paradox of Optimal Repentance](#)

*"When they inquire into predestination, they are penetrating the sacred precincts of divine wisdom. If anyone with carefree assurance breaks into this place, he will not succeed in satisfying his curiosity and he will enter a labyrinth from which he can find no exit."*

-- John Calvin

John Calvin preached the doctrine of predestination: that God irreversibly decreed each man's eternal fate at the moment of Creation. Calvinists separate mankind into two groups: the elect, whom God predestined for Heaven, and the reprobate, whom God predestined for eternal punishment in Hell.

If you had the bad luck to be born a sinner, there is nothing you can do about it. You are too corrupted by original sin to even have the slightest urge to seek out the true faith. Conversely, if you were born one of the elect, you've got it pretty good; no matter what your actions on Earth, it is impossible for God to revoke your birthright to eternal bliss.

However, it is believed that the elect always live pious, virtuous lives full of faith and hard work. Also, the reprobate always commit heinous sins like greed and sloth and commenting on anti-theist blogs. This isn't what causes God to damn them. It's just what happens to them after they've been damned: their soul has no connection with God and so it tends in the opposite direction.

Consider two Calvinists, Aaron and Zachary, both interested only in maximizing his own happiness. Aaron thinks to himself "Whether or not I go to Heaven has already been decided, regardless of my actions on Earth. Therefore, I might as well try to have as much fun as possible, knowing it won't effect the afterlife either way." He spends his days in sex, debauchery, and anti-theist blog comments.

Zachary sees Aaron and thinks "That sinful man is thus proven one of the reprobate, and damned to Hell. I will avoid his fate by living a pious life." Zachary becomes a great minister, famous for his virtue, and when he dies his entire congregation concludes he must have been one of the elect.

Before the cut: If you were a Calvinist, which path would you take?

Amos Tversky, Stanford psychology professor by day, [bias-fighting superhero by night](#), thinks you should live a life of sin. He bases [his analysis](#) of the issue on the famous maxim that [correlation is not causation](#). Your virtue during life is correlated to your eternal reward, but only because they're both correlated to a hidden [third variable](#), your status as one of the elect, which causes both.

Just to make that more concrete: people who own more cars live longer. Why? Rich people buy more cars, and rich people have higher life expectancies. Both cars and long life are caused by a hidden third variable, wealth. Trying to increase your chances

of getting into Heaven by being virtuous is as futile as trying to increase your life expectancy by buying another car.

Some people would stop there, but not Amos Tversky, bias-fighting superhero. He and George Quattrone conducted [a study](#) that both illuminated a flaw in human reasoning about causation and demonstrated yet another way people can be simultaneously right and wrong.

Subjects came in thinking it was a study on cardiovascular health. First, experimenters tested their pain tolerance by making them stick their hands in a bucket of freezing water until they couldn't bear it any longer. However long they kept it there was their baseline pain tolerance score.

Then experimenters described two supposed types of human heart: Type I hearts, which work poorly and are prone to heart attack and will kill you at a young age, and Type II hearts, which work well and will bless you with a long life. You can tell a Type I heart from a Type II heart because...and here the subjects split into two groups. Group A learned that people with Type II hearts, the good hearts, had higher pain tolerance after exercise. Group B learned that Type II hearts had lower pain tolerance after exercise.

Then the subjects exercised for a while and stuck their hands in the bucket of ice water again. Sure enough, the subjects who thought increased pain tolerance meant a healthier heart kept their hands in longer. And then when the researchers went and asked them, they said they must have a Type II heart because the ice water test went so well!

The subjects seem to have believed *on some level* that keeping their hand in the water longer could give them a different kind of heart. Dr. Tversky declared that people have a cognitive blind spot to "hidden variable" causation, and this explains the Calvinists who made such an effort to live virtuously.

But this study is also interesting as an example of self-deception. One level of the mind made the (irrational) choice to leave the hand in the ice water longer. Another level of the mind that wasn't consciously aware of this choice interpreted it as evidence for the Type II heart. There are two cognitive flaws here: the subject's choice to try harder on the ice water test, and his lack of realization that he'd done so.

I don't know of any literature explicitly connecting this study to [self-handicapping](#), but the surface similarities are striking. In both, a person takes an action intended to protect his self-image that will work if and only if he doesn't realize this intent. In both, the action is apparently successful, self-image is protected, and the conscious mind remains unaware of the true motives.

Despite all this, and with all due respect to Dr. Tversky I think he might be wrong about the whole predestination issue. If I were a Calvinist, I'd live a life of sin if and only if I would two-box on [Newcomb's Problem](#).

# Talking Snakes: A Cautionary Tale

I particularly remember one scene from Bill Maher's "[Religulous](#)". I can't find the exact quote, but I will try to sum up his argument as best I remember.

Christians believe that sin is caused by a talking snake. They may have billions of believers, thousands of years of tradition behind them, and a vast literature of apologetics justifying their faith - but when all is said and done, they're adults who believe in a talking snake.

I have read of the absurdity heuristic. I know that it is not *carte blanche* to go around rejecting beliefs that seem silly. But I was still sympathetic to the talking snake argument. After all...a *talking snake*?

I changed my mind in a Cairo cafe, talking to a young Muslim woman. I let it slip during the conversation that I was an atheist, and she seemed genuinely curious why. You've all probably been in such a situation, and you probably know how hard it is to choose just one reason, but I'd been reading about Biblical contradictions at the time and I mentioned the myriad errors and atrocities and contradictions in all the Holy Books.

Her response? "Oh, thank goodness it's that. I was afraid you were one of those crazies who believed that monkeys transformed into humans."

I admitted that [um, well, maybe I sorta kinda](#) might in fact believe that.

It is hard for me to describe exactly the look of shock on her face, but I have no doubt that her horror was genuine. I may have been the first flesh-and-blood evolutionist she ever met. "But..." she looked at me as if I was an idiot. "Monkeys don't change into humans. What on Earth makes you think monkeys can change into humans?"

I admitted that the whole process was rather complicated. I suggested that it wasn't exactly a Optimus Prime-style transformation so much as a gradual change over eons and eons. I recommended a few books on evolution that might explain it better than I could.

She said that she respected me as a person but that quite frankly I could save my breath because there was no way any book could possibly convince her that monkeys have human babies or whatever sort of balderdash I was preaching. She accused me and other evolution believers of being too willing to accept absurdities, motivated by our atheism and our fear of the self-esteem hit we'd take by accepting Allah was greater than ourselves.

It is not clear to me that this woman did anything differently than Bill Maher. Both heard statements that sounded so crazy as to not even merit further argument. Both recognized that there was a large group of people who found these statements plausible and had written extensive literature justifying them. Both decided that the statements were so absurd as to not merit examining that literature more closely. Both came up with reasons why they could discount the large number of believers because those believers must be biased.

I post this as a cautionary tale as we [discuss the logic](#) or [illogic](#) of theism. I propose

taking from it the following lessons:

- The [absurdity heuristic](#) doesn't work very well.
- Even on [things that sound really, really absurd](#).
- If a large number of intelligent people believe something, it deserves your attention. After you've studied it on its own terms, then you have a right to reject it. You could still be wrong, though.
- Even if you can think of a good reason why people might be biased towards the silly idea, thus explaining it away, your good reason may still be false.
- If someone cannot explain why something is not stupid to you over twenty minutes at a cafe, that doesn't mean it's stupid. It just means it's complicated, or they're not very good at explaining things.
- There is no royal road.

*(special note to those prone to [fundamental attribution errors](#): I do not accept theism. I think theism is wrong. I think it can be demonstrated to be wrong on logical grounds. I think the nonexistence of talking snakes is evidence against theism and can be worked into a general argument against theism. I just don't think it's as easy as saying "talking snakes are silly, therefore theism is false." And I find it embarrassing when atheists say things like that, and then get called on it by intelligent religious people.)*

# The Skeptic's Trilemma

**Followup to:** [Talking Snakes: A Cautionary Tale](#)

**Related to:** [Explain, Worship, Ignore](#)

Skepticism is like sex and pizza: when it's good, it's very very good, and when it's bad, it's still pretty good.

It really is hard to dislike skeptics. Whether or not their rational justifications are perfect, they are doing society a service by raising the social cost of holding false beliefs. But there is a failure mode for skepticism. It's the same as the failure mode for so many other things: it becomes a [blue vs. green](#) style tribe, [demands support](#) of all 'friendly' arguments, enters an [affective death spiral](#), and collapses into a cult.

What does it look like when skepticism becomes a cult? Skeptics become more interested in supporting their "team" and insulting the "enemy" than in finding the truth or convincing others. They begin to think "If assigning .001% probability to Atlantis and not accepting its existence without extraordinarily compelling evidence is good, then assigning 0% probability to Atlantis and refusing to even consider any evidence for its existence must be *great*!" They begin to deny any evidence that seems pro-Atlantis, and cast aspersions on the character of anyone who produces it. They become anti-Atlantis fanatics.

Wait a second. There is no lost continent of Atlantis. How do I know what a skeptic would do when confronted with evidence for it? For that matter, why do I care?

Way back in 2007, Eliezer described the rationalist equivalent of Abort, Retry, Fail: the trilemma of [Explain, Worship, Ignore](#). Don't understand where rain comes from? You can try to explain it as part of the water cycle, although it might take a while. You can worship it as the sacred mystery of the rain god. Or you can ignore it and go on with your everyday life.

So someone tells you that Plato, normally a pretty smart guy, wrote [a long account](#) of a lost continent called Atlantis complete with a bunch of really specific geographic details that seem a bit excessive for a meaningless allegory. Plato claims to have gotten most of the details from a guy called [Solon](#), legendary for his honesty, who got them from the Egyptians, who are known for their obsessive record-keeping. This seems interesting. But there's no evidence for a lost continent anywhere near the Atlantic Ocean, and geology tells us continents can't just go missing.

One option is to hit Worship. Between the Theosophists, Edgar Cayce, the Nazis, and a bunch of well-intentioned but crazy amateurs including [a U.S. Congressman](#), we get a supercontinent with technology far beyond our wildest dreams, littered with glowing crystal pyramids and powered by the peaceful and eco-friendly mystical wisdom of the ancients, source of all modern civilization and destined to rise again to herald the dawning of the Age of Aquarius.

Or you could hit Ignore. I accuse the less pleasant variety of skeptic of taking this option. Atlantis is stupid. Anyone who believes it is stupid. Plato was a dirty rotten liar. Any scientist who finds anomalous historical evidence suggesting a missing piece to the early history of the Mediterranean region is also a dirty rotten liar, motivated by



crazy New Age beliefs, and should be fired. Anyone who talks about Atlantis is the Enemy, and anyone who denies Atlantis gains immediate access to our in-group and official Good Rational Scientific Person status.

[Spyridon Marinatos](#), a Greek archaeologist who really deserves more fame than he received, was a man who hit Explain. The geography of Plato's Atlantis, a series of concentric circles of land and sea, had been derided as fanciful; Marinatos noted<sup>1</sup> that it matched the geography of the Mediterranean island of Santorini quite closely. He also noted that Santorini had a big volcano right in the middle and seemed somehow linked to the Minoan civilization, a glorious race of seafarers who had mysteriously collapsed a thousand years before Plato. So he decided to go digging in Santorini. And he found...

...the lost city of Atlantis. Well, I'm making an assumption here. But [the city he found](#) was over four thousand years old, had a population of over ten thousand people at its peak, boasted three-story buildings and astounding works of art, and had hot and cold running water - an unheard-of convenience that it shared with the city in Plato's story. For the Early Bronze Age, that's *darned* impressive. And like Plato's Atlantis, it was destroyed in a single day. The volcano that loomed only a few miles from its center [went off around 1600 BC](#), utterly burying it and destroying its associated civilization. No one knows what happened to the survivors, but the most popular theory is that some fled to Egypt<sup>2</sup>, with which the city had flourishing trade routes at its peak.

The Atlantis = Santorini equivalence is still controversial, and the point of this post isn't to advocate for it. But just look at the difference between Joe Q. Skeptic and Dr. Marinatos. Both were rightly skeptical of the crystal pyramid story erected by the Atlantis-worshippers. But Joe Q. Skeptic considered the whole issue a nuisance, or at best a way of proving his intellectual superiority over the believers. Dr. Marinatos saw an honest mystery, developed a theory that made testable predictions, then went out and started digging.

The fanatical skeptic, when confronted with some evidence for a seemingly paranormal claim, says "Wow, that's stupid." It's a soldier on the opposing side, and the only thing to be done with it is kill it as quickly as possible. The wise skeptic, when confronted with the same evidence, says "Hmmm, that's interesting."

Did people at Roswell discovered the debris of a strange craft made of seemingly otherworldly material lying in a field, only to be silenced by the government later? You can worship the mighty aliens who are cosmic bringers of peace. You can ignore it, because UFOs don't exist so the people are clearly lying. Or you can search for an explanation until you find that the government was conducting tests of [Project Mogul](#) in that very spot.

Do thousands of people claim that therapies with no scientific basis are working? You can worship alternative medicine as a natural and holistic alternative to stupid evil materialism. You can ignore all the evidence for their effectiveness. Or you can shut up and discover the placebo effect, explaining the lot of them in one fell swoop.

Does someone claim to see tiny people, perhaps elves, running around and doing elvish things? You can call them lares and worship them as household deities. You can ignore the person because he's an obvious crank. Or you can go to a neurologist, and he'll explain that the person's probably suffering from [Charles Bonnet](#) Syndrome.

All unexplained phenomena are real. That is, they're real unexplained phenomena. The explanation may be prosaic, like that people are gullible. Or it may be an entire four thousand year old lost city of astounding sophistication. But even "people are gullible" can be an interesting explanation if you're smart enough to make it one. There's a big difference between "people are gullible, so they believe in stupid things like religion, let's move on" and a complete list of the cognitive biases that make explanations involving agency and intention more attractive than naturalistic explanations to a naive human mind. A [sufficiently intelligent](#) thinker could probably reason from the mere existence of religion all the way back to the fundamentals of evolutionary psychology.

This I consider a specific application of a more general rationalist technique: not prematurely dismissing things that go against your worldview. There's a big difference between dismissing that whole Lost Continent of Atlantis story, and *prematurely* dismissing it. It's the difference between discovering an ancient city and resting smugly satisfied that you don't have to.

## Footnotes

**1:** I may be unintentionally sexing up the story here. I read a book on Dr. Marinatos a few years ago, and I know he did make the Santorini-Atlantis connection, but I don't remember whether he made it before starting his excavation, or whether it only clicked during the dig (and the Internet is silent on the matter). If it was the latter, all of my moralizing about how wonderful it was that he made a testable prediction falls a bit flat. I should have used another example where I knew for sure, but this story was too perfect. Mea culpa.

**2:** I don't include it in the main article because it is highly controversial and you have to fudge some dates for it to really work out, but here is a Special Bonus Scientific Explanation of a Paranormal Claim: the eruption of this same supervolcano in 1600 BC [caused](#) the series of geologic and climatological catastrophes recorded in the Bible as the Ten Plagues of Egypt. However, I specify that I'm including this because it's fun to think about rather than because there's an especially large amount of evidence for it.

# Are You a Solar Deity?

Max Muller was one of the greatest religious scholars of the 19th century. Born in Germany, he became fascinated with Eastern religion, and moved to England to be closer to the center of Indian scholarship in Europe. There he mastered English and Sanskrit alike to come out with the first English translation of the Rig Veda, the holiest book of Hinduism.

One of Muller's most controversial projects was his attempt to interpret all pagan mythologies as linked to one another, deriving from a common ur-mythology and ultimately from the celestial cycle. His tools were exhaustive knowledge of the myths of all European cultures combined with a belief in the interpretive power of linguistics.

What the [significance](#) of Orpheus' descent into the underworld to reclaim his wife's soul? The sun sets beneath the Earth each evening, and returns with renewed brightness. Why does Apollo love Daphne? Daphne is cognate with Sanskrit Dahana, the maiden of the dawn. The death of Hercules? It occurs after he's completed twelve labors (cf. twelve signs of zodiac) when he's travelling west (like the sun), he is killed by Deianeira (compare Sanskrit dasya-nari, a demon of darkness) and his body is cremated (fire = the sun). His followers extended the method to Jesus - who was clearly based on a lunar deity, since he spent three days dead and then returned to life, just as the new moon goes dark for three days and then reappears.

Muller's work was massively influential during his time, and many 19th century mythographers tried to critique his paradigm and poke holes in it. Some accused him of trying to destroy the mystery of religion, and others accused him of shoddy scholarship.

R.F. Littledale, an Anglican apologist, took a completely different route. He claimed that there was, in fact, no such person as Professor Max Muller, holder of the Taylorian Chair in Modern European Languages. All these stories about "Max Muller" were nothing but a thinly disguised solar myth.

Littledale begins his argument by noting Muller's heritage. He was supposedly born in Germany, only to travel to England when he came of age. This looks suspiciously like the classic Journey of the Sun, which is born in the east but travels to the west. Muller's origin in Germany is a clear reference to Germanus Apollo, one of the old appellations of the Greek sun god.

His Christian name must be related to Latin "maximus" or Sanskrit "maha", meaning great, a suitable description of the King of Gods, and his surname is cognate with Mjolnir, the mighty hammer of the sky god Thor. His claim to fame is bringing the ancient wisdom of the East to the people of the West - that is, illuminating them with eastern light.

Muller teaches at Oxford for the same reason that Genesis describes the sky as "the waters above" and the Egyptians gave Ra a solar barge: ancient people interpreted the sky as a river, and the sun as crossing that river upon his chariot (perhaps an **ox**-drawn chariot, **fording** the river?). His chair at Oxford is the throne of the sky, his status as Taylorian Professor because "he cuts away with his glittering shears the ragged edges of cloud; he allows the...cuttings from his workshop, to descend in fertilizing showers upon the earth."

I could go on; instead I recommend you read [the original essay](#). The take-home lesson is that any technique powerful enough to prove that Hercules is a solar myth is also powerful enough to prove that *anyone* is a solar myth. Muller lacked the [strength of a rationalist](#): the ability to be more confused by fiction than by reality. This makes the Hercules theory useless, but that is not immediately apparent on a first or even a second reading of Muller's work. When reading Muller's work, the primary impression one gets is "Wow, this man has gathered a *lot* of supporting evidence."

This is a problem encountered in many fields of scholarship, especially "comparative" anything. In comparative linguistics, for example, it's usually possible to make a [case that two languages are related](#) good enough to convince a layman, no matter which two languages or how distant they may be. In comparative religion, we get cases like this blog's recent discussion over the [possible derivation of Esther and Mordechai defeating Haman](#) from Ishtar and Marduk defeating Humbaba. The less said about comparative literature, the better, although I can't help but quote humor writer Dave Barry:

Suppose you are studying Moby-Dick. Anybody with any common sense would say that Moby-Dick is a big white whale, since the characters in the book refer to it as a big white whale roughly eleven thousand times. So in *your* paper, *you* say Moby-Dick is actually the Republic of Ireland. Your professor, who is sick to death of reading papers and never liked Moby-Dick anyway, will think you are enormously creative. If you can regularly come up with lunatic interpretations of simple stories, you should major in English.

The worst (but most fun to read!) are in pseudoscience, where plausible sounding comparisons can prove almost anything. Did you know the Mayans believed in a lost homeland called Atzlan, the Indonesians believed in a lost island called Atala, and the Greeks believed in a lost continent called Atlantis? Likewise, did you know that Nostradamus predicted a great battle involving Germany and "Hister", which sounds almost like "Hitler"?

Yet it would be a mistake to reject all such comparisons. In fact, I have thus far been enormously unfair to Professor Muller, whose work established several correspondences still viewed as valid today. Virtually all modern mythologists accept that the Hindu Varuna is the Greek Uranus, and that the Greek sky god Zeus equals the Hindu sky god Dyaus Pita and the Roman Jupiter (compare to Latin *deus pater*, meaning God the Father). Likewise, comparative linguists are quite certain that all modern European languages and Sanskrit derive from a common Indo-European root, and in my opinion even the Nostratic project - an ambitious attempt to link Semitic, Indo-European, Uralic, and a bunch of other languages - is at least worth consideration.

We need a test to distinguish between true and false correspondences. But the standard method, making and testing predictions, is useless here. A good mythologist already knows the stories of Varuna and Uranus. The chances of discovering a new fact that either confirms or overturns the Varuna-Uranus correspondence is not even worth considering.

Mark Rosenfelder has [an excellent article on chance resemblances between languages](#) which offers a semi-formal model for spotting dubious comparisons. But such precision may not be possible when comparing two deities.

I have what might be a general strategy for approaching this sort of problem, which I will present tomorrow. But how would you go about it?

# The "Spot the Fakes" Test

**Followup to:** [Are You a Solar Deity?](#)

James McAuley and Harold Stewart were mid-20th century Australian poets, and they were not happy. After having society ignore their poetry in favor of "experimental" styles they considered fashionable nonsense, they wanted to show everyone what they already knew: the Australian literary world was full of empty poseurs.

They began by selecting random phrases from random books. Then they linked them together into something sort of like poetry. Then they invented the most fashionable possible story: Ern Malley, a loner working a thankless job as an insurance salesman, writing sad poetry in his spare time and hiding it away until his death at an early age. Posing as Malley's sister, who had recently discovered the hidden collection, they sent the works to *Angry Penguins*, one of Australia's top experimental poetry magazines.

You wouldn't be reading this if the magazine hadn't rushed a special issue to print in honor of "a poet in the same class as W.H. Auden or Dylan Thomas".

[The hoax](#) was later revealed<sup>1</sup>, everyone involved ended up with egg on their faces, and modernism in Australia received a serious blow. But as I am reminded every time I look through a modern poetry anthology, one Ern Malley every fifty years just isn't enough. I daydream about an alternate dimension where people are genuinely interested in keeping literary criticism honest. In this universe, any would-be literary critic would have to distinguish between ten poems generally recognized as brilliant that he'd never seen before, and ten pieces of nonsense invented on the spot by drunk college students, in order to keep his critic's license.

Can we refine this test? And could it help Max Muller with his solar deity problem?

In the Malley hoax, McAuley and Stewart suspected that a certain school of modernist poetry was without value. Because its supporters were too biased to admit this directly, they submitted a control poem they knew was without value, and found the modernists couldn't tell the difference. This suggests a powerful technique for determining when something otherwise untestable might be, as Neal Stephenson calls it, *bulshytte*.

Perhaps Max Muller thinks Hercules is a solar deity. He will write up a argument for this proposition, and submit it for consideration before all the great mythologists of the world. Even if these mythologists want to be unbiased, they will have a difficult time of it: Muller has a prestigious reputation, and they may not have any set conception of what does and doesn't qualify as a solar deity.

What if, instead of submitting one argument, Muller submitted ten? One sincere argument for why Hercules is a solar deity, and other bogus arguments for why Perseus, Bellerophon, Theseus, et cetera are solar myths (which he has nevertheless constructs to the best of his ability). Then he instructs the mythologists "Please independently determine which of these arguments is true, and which ones I have just come up with by writing 'X is a solar deity' as my bottom line and then inventing fake justifications for the fact?" If every mythologist finds the Hercules argument most convincing, then that doesn't prove anything about Hercules but it at least shows Muller has a strong case. On the other hand, if they're all convinced by different



arguments, or find none of the arguments convincing, or worst of all they all settle on Bellerophon, then Dr. Muller knows his beliefs about Hercules are quite probably wishful thinking.

This method hinges on Dr. Muller's personal honesty: a dishonest man could simply do a bad job arguing for Theseus and Bellerophon. What if we thought Dr. Muller was dishonest? We might find another mythologist whom independent observers rate as equally persuasive as Dr. Muller, and ask her to come up with the bogus arguments.

The rationalists I know sometimes take a dim view of the humanities as academic disciplines. Part of the problem is the seeming untestability of their conclusions through good, blinded experimental methods. I don't think most humanities professors are really looking all that hard for such methods. But for those who are, I consider this technique a little better than nothing<sup>2</sup>.

### Footnotes

**1:** The [Sokal Affair](#) is another related hoax. Wikipedia's Sokal Hoax page has [some other excellent examples](#) of this sort of test.

**2:** One more example where this method could prove useful. I remember debating a very smart Christian on the subject of Biblical atrocities. You know, stuff about death by stoning for minor crimes, or God ordering the Israelites to murder women and enslave children - that sort of thing. My friend, who was quite smart, was always able to come up with a superficially plausible excuse, and it was getting on my nerves. But having just read [Your Strength as a Rationalist](#), I knew that being able to explain anything wasn't always a virtue. I proposed the following experiment: I'd give my friend ten atrocities commanded by random Bronze Age kings generally agreed by historical consensus to be jerks, and ten commanded by God in the Bible. His job would be to determine which ten, for whatever reason, really weren't all that bad. If he identified the ten Bible passages, that would be strong evidence that Biblical commandments only seemed atrocious when misunderstood. But if he couldn't tell the difference between God and Ashurbanipal, that would prove God wasn't really that great. To my disgust, my friend knew his Bible so well that I couldn't find any atrocities he wasn't already familiar with. So much for that technique. I offer it to anyone who debates theists with less comprehensive knowledge of Scripture.

# How to Not Lose an Argument

**Related to:** [Leave a Line of Retreat](#)

**Followup to:** [Talking Snakes: A Cautionary Tale](#), [The Skeptic's Trilemma](#)

*"I argue very well. Ask any of my remaining friends. I can win an argument on any topic, against any opponent. People know this, and steer clear of me at parties. Often, as a sign of their great respect, they don't even invite me."*

--Dave Barry

The science of winning arguments is called Rhetoric, and it is one of the Dark Arts. Its study is forbidden to rationalists, and its tomes and treatises are kept under lock and key in a particularly dark corner of the Miskatonic University library. More than this it is not lawful to speak.

But I do want to talk about a very closely related skill: not losing arguments.

Rationalists probably find themselves in more arguments than the average person. And if we're doing it right, the truth is hopefully on our side and the argument is ours to lose. And far too often, we *do* lose arguments, even when we're right. Sometimes it's because of biases or inferential distances or other things that can't be helped. But all too often it's because we're shooting ourselves in the foot.

How does one avoid shooting one's self in the foot? In rationalist language, the technique is called Leaving a Social Line of Retreat. In normal language, it's called being nice.

First, what does it mean to win or lose an argument? There is an unspoken belief in some quarters that the point of an argument is to gain social status by utterly demolishing your opponent's position, thus proving yourself the better thinker. That can be fun sometimes, and if it's really all you want, go for it.

But the most important reason to argue with someone is to change his mind. If you want a world without fundamentalist religion, you're never going to get there just by making cutting and incisive critiques of fundamentalism that all your friends agree sound really smart. You've got to deconvert some actual fundamentalists. In the absence of changing someone's mind, you can at least get them to see your point of view. Getting fundamentalists to understand the [real reasons](#) people find atheism attractive is a nice consolation prize.

I make the anecdotal observation that a lot of smart people are very good at winning arguments in the first sense, and very bad at winning arguments in the second sense. Does that correspond to your experience?

Back in 2008, Eliezer described how to [Leave a Line of Retreat](#). If you believe morality is impossible without God, you have a strong disincentive to become an atheist. Even after you've realized which way the evidence points, you'll activate every possible defense mechanism for your religious beliefs. If all the defense mechanisms fail, you'll take God on utter faith or just [believe in belief](#), rather than surrender to the unbearable position of an immoral universe.

The correct procedure for dealing with such a person, Eliezer suggests, isn't to show them yet another reason why God doesn't exist. They'll just reject it along with all the others. The correct procedure is to convince them, on a gut level, that morality is possible even in a godless universe. When disbelief in God is no longer so terrifying, people won't fight it quite so hard and may even deconvert themselves.

But there's another line of retreat to worry about, one I experienced firsthand in a very strange way. I had a dream once where God came down to Earth; I can't remember exactly why. In the borderlands between waking and sleep, I remember thinking: *I feel like a total moron*. Here I am, someone who goes to atheist groups and posts on atheist blogs and has told all his friends they should be atheists and so on, and now it turns out God exists. All of my religious friends whom I won all those arguments against are going to be secretly looking at me, trying as hard as they can to be nice and understanding, but secretly laughing about how I got my comeuppance. I can never show my face in public again. Wouldn't you feel the same?

And then I woke up, and shook it off. I am an aspiring rationalist: if God existed, I would desire to believe that God existed. But I realized at that point the importance of the social line of retreat. The psychological resistance I felt to admitting God's existence, even after having seen Him descend to Earth, was immense. And, I realized, it was exactly the amount of resistance that every vocally religious person must experience towards God's *non*-existence.

There's not much we can do about this sort of high-grade long-term resistance. Either a person has enough of the [rationalist virtues](#) to overcome it, or he doesn't. But there is a less ingrained, more immediate form of social resistance generated with every heated discussion.

Let's say you approach a theist (let's call him Theo) and say "How can you, a grown man, still believe in something stupid like [talking snakes](#) and magic sky kings? Don't you know you people are responsible for the Crusades and the Thirty Years' War and the Spanish Inquisition? You should be ashamed of yourself!"

This suggests the following dichotomy in Theo's mind: **EITHER** God exists, **OR** I am an idiot who believes in stupid childish things and am in some way partly responsible for millions of deaths and I should have lower status and this arrogant person who's just accosted me and whom I already hate should have higher status at my expense.

Unless Theo has attained a level of rationality far beyond any of us, guess which side of that dichotomy he's going to choose? In fact, guess which side of that dichotomy he's now going to support with renewed vigor, even if he was only a lukewarm theist before? His social line of retreat has been completely closed off, and it's *your* fault.

Here the two definitions of "winning an argument" I suggested before come into conflict. If your goal is to absolutely demolish the other person's position, to make him feel awful and worthless - then you are also very unlikely to change his mind or win his understanding. And because our culture of debates and mock trials and real trials and flaming people on Usenet encourages the first type of "winning an argument", there's precious little genuine mind-changing going on.

Really adjusting to the second type of argument, where you try to convince people, takes a lot more than just not insulting people outright<sup>1</sup>. You've got to completely

rethink your entire strategy. For example, anyone used to the Standard Debates may already have a cached pattern of how they work. Activate the whole Standard Debate concept, and you activate a whole bunch of related thoughts like Atheists As The Enemy, Defending The Faith, and even in some cases (I've seen it happen) persecution of Christians by atheists in Communist Russia. To such a person, ceding an inch of ground in a Standard Debate may well be equivalent to saying all the Christians martyred by the Communists died in vain, or something similarly dreadful.

So try to show you're not just starting Standard Debate #4457. I remember once, during the middle of a discussion with a Christian, when I admitted I really didn't like Christopher Hitchens. Richard Dawkins, brilliant. Daniel Dennett, brilliant. But Christopher Hitchens always struck me as too black-and-white and just plain irritating. This one little revelation completely changed the entire tone of the conversation. I was no longer Angry Nonbeliever #116. I was no longer the living incarnation of All Things Atheist. I was just a person who happened to have a whole bunch of atheist ideas, along with a couple of ideas that weren't typical of atheists. I got the same sort of response by admitting I loved religious music. All of a sudden my friend was falling over himself to mention some scientific theory he found especially elegant in order to reciprocate<sup>2</sup>. I didn't end up deconverting him on the spot, but think he left with a much better appreciation of my position.

All of these techniques fall dangerously close to the Dark Arts, so let me be clear: I'm not suggesting you misrepresent yourself just to win arguments. I don't think misrepresenting yourself would even work; evolutionary psychology tells us humans are notoriously bad liars. Don't fake an appreciation for the other person's point of view, actually *develop* an appreciation for the other person's point of view. Realize that [your points probably seem as absurd to others as their points seem to you](#). Understand that [many false beliefs don't come from simple lying or stupidity](#), but from complex mixtures of truth and falsehood filtered by complex cognitive biases. Don't stop believing that you are right and they are wrong, unless the evidence points that way. But leave it at them being wrong, not them being wrong and stupid and evil.

I think most people intuitively understand this. But considering how many smart people I see shooting their own foot off when they're trying to convince someone<sup>3</sup>, some of them clearly need a reminder.

## Footnotes

**1:** An excellent collection of the deeper and most subtle forms of this practice of this sort can be found in Dale Carnegie's *How to Win Friends and Influence People*, one of the only self-help books I've read that was truly useful and not a regurgitation of clichés and applause lights. Carnegie's thesis is basically that being nice is the most powerful of the Dark Arts, and that a master of the Art of Niceness can use it to take over the world. It works better than you'd think.

**2:** The following technique is definitely one of the Dark Arts, but I mention it because it reveals a lot about the way we think: when engaged in a really heated, angry debate, one where the insults are flying, suddenly stop and admit the other person is one hundred percent right and you're sorry for not realizing it earlier. Do it properly, and the other person will be flabbergasted, and feel deeply guilty at all the names and bad feelings they piled on top of you. Not only will you ruin their whole day, but for

the rest of time, this person will secretly feel indebted to you, and you will be able to play with their mind in all sorts of little ways.

**3:** Libertarians, you have a particular problem with this. If I wanted to know why I'm a Stalin-worshipper who has betrayed the Founding Fathers for personal gain and is controlled by his base emotions and wants to dominate others by force to hide his own worthlessness et cetera, I'd ask Ann Coulter. You're better than that. Come on. And then you wonder why people never vote for you.

# The Power of Positivist Thinking

**Related to:** [No Logical Positivist I](#), [Making Beliefs Pay Rent](#), [How An Algorithm Feels From Inside](#), [Disguised Queries](#)

Call me non-conformist, call me one man against the world, but...I kinda like logical positivism.

The logical positivists were a dour, no-nonsense group of early 20th-century European philosophers. Indeed, the phrase "no-nonsense" seems almost invented to describe the Positivists. They liked nothing better than to reject the pet topics of other philosophers as being untestable and therefore meaningless. Is the true also the beautiful? Meaningless! Is there a destiny to the affairs of humankind? Meaningless? What is justice? Meaningless! Are rights inalienable? Meaningless!

Positivism became stricter and stricter, defining more and more things as meaningless, until someone finally pointed out that positivism itself was meaningless by the positivists' definitions, at which point the entire system vanished in a puff of logic. Okay, it wasn't that simple. It took several decades and Popper's falsifiability to seal its coffin. But vanish it did. It remains one of the least lamented theories in the history of philosophy, because if there is one thing philosophers hate it's people telling them they can't argue about meaningless stuff.

But if we've learned anything from fantasy books, it is that any cabal of ancient wise men destroyed by their own hubris at the height of their glory must leave behind a single ridiculously powerful artifact, which in the right hands gains the power to dispel darkness and annihilate the forces of evil.

The positivists left us the idea of verifiability, and it's time we started using it more.

Eliezer, in [No Logical Positivist I](#), condemns the positivist notion of verifiability for excluding some perfectly meaningful propositions. For example, he says, it may be that a chocolate cake formed in the center of the sun on 8/1/2008, then disappeared after one second. This statement seems to be meaningful; that is, there seems to be a difference between it being true or false. But there's no way to test it (at least without time machines and sundiver ships, which we can't prove are possible) so the logical positivists would dismiss it as nonsense.

I am not an expert in logical positivism; I have two weeks studying positivism in an undergrad philosophy class under my belt, and little more. If Eliezer says that is how the positivists interpreted their verifiability criterion, I believe him. But it's not the way I would have done things, if I'd been in 1930s Vienna. I would have said that any statement corresponding to a state of the material universe, reducible in theory to things like quarks and photons, testable by a being who has access to the machine running the universe<sup>1</sup> and who can check the logs at will - such a statement is meaningful<sup>2</sup>. In this case the chocolate cake example passes: it corresponds to a state of the material world, and is clearly visible on the universe's logs. "Rights are inalienable" remains meaningless, however. At the risk of reinventing the wheel<sup>3</sup>, I will call this interpretation "soft positivism".



My positivism gets even softer, though. Consider the statement "Google is a successful company." Though my knowledge of positivism is shaky, I believe that most positivists would reject this as meaningless; "success" is too fuzzy to be reduced to anything objective. But if positivism is true, it should add up to normality: we shouldn't find that an obviously useful statement like "Google is a successful company" is total nonsense. I interpret the statement to mean certain objectively true propositions like "The average yearly growth rate for Google has been greater than the average yearly growth rate for the average company", which itself reduces down to a question of how much money Google made each year, which is something that can be easily and objectively determined by anyone with the universe's logs.

I'm not claiming that "Google is a successful company" has an absolute one-to-one identity with a statement about average growth rates. But the "successful company" statement is clearly allied with many testable statements. Average growth rate, average profits per year, change in the net worth of its founders, numbers of employees, et cetera. Two people arguing about whether Google was a successful company could in theory agree to create a formula that captures as much as possible of their own meaning of the word "successful", apply that formula to Google, and see whether it passed. To say "Google is a successful company" reduces to "I'll bet if we established a test for success, which we are not going to do, Google would pass it."

(Compare this to [Eliezer's meta-ethics](#), where he says "X is good" reduces to "I'll bet if we calculated out this gigantic human morality computation, which we are not going to do, X would satisfy it.")

This can be a very powerful method for resolving debates. I remember getting into an argument with my uncle, who believed that Obama's election would hurt America because having a Democratic president is bad for the economy. We were doing the normal back and forth, him saying that Democrats raised taxes which discouraged growth, me saying that Democrats tended to be more economically responsible and less ideologically driven, and we both gave lots of examples and we never would have gotten anywhere if I hadn't said "You know what? Can we both agree that this whole thing is basically asking whether average GDP is lower under Democratic than Republican presidents?" And he said "Yes, that's pretty much what we're arguing about." So I went and got [the GDP statistics](#), sure enough they were higher under Democrats, and he admitted I had a point<sup>4</sup>.

But people aren't always as responsible as my uncle, and debates aren't always reducible to anything as simple as GDP. Consider: Zahra approaches Aaron and says: "Islam is a religion of peace."<sup>5</sup>

Perhaps Aaron disagrees with this statement. Perhaps he begins debating. There are many things he could say. He could recall all the instances of Islamic terrorism, he could recite seemingly violent verses from the Quran, he could appeal to wars throughout history that have involved Muslims. I've heard people try all of these.

And Zahra will respond to Aaron in the same vein. She will recite Quranic verses praising peace, and talk about all the peaceful Muslims who never engage in terrorism at all, and all of the wars started by Christians in which Muslims were innocent victims. I have heard all these too.

Then Paula the Positivist comes by. "Hey," she says, "We should reduce this statement to testable propositions, and then there will be no room for disagreement."

But maybe, if asked to estimate the percentage of Muslims who are active in terrorist groups, Aaron and Zahra will give the exact same number. Perhaps they are both equally aware of all the wars in history in which Muslims were either aggressors or peacemakers. They may both have the entire Quran memorized and be fully aware of all appropriate verses. But even after Paula has checked to make sure they agree on every actual real world fact, there is no guarantee that they will agree on whether Islam is a religion of peace or not.

What if we ask Aaron and Zahra to reduce "Islam is a religion of peace" to an empirical proposition? In the best case, they will agree on something easy, like "Muslims on average don't commit any more violent crimes than non-Muslims." Then you just go find some crime statistics and the problem is solved. In the second-best case, the two of them reduce it to completely different statements, like "No Muslim has ever committed a violent act" versus "Not all Muslims are violent people." This is still a resolution to the argument; both Aaron and Zahra may agree that the first proposition is false and the second proposition is true, and they both agree the original statement was too vague to go around professing.

In the worst-case scenario, they refuse to reduce the statement at all, or they deliberately reduce it to something untestable, or they reduce it to two different propositions but are outraged that their opponent is using a different proposition than they are and think their opponent's proposition is clearly not equivalent to the original statement.

How are they continuing to disagree, when they agree on all of the relevant empirical facts and they fully understand the concept of reducing a proposition?

In [How an Algorithm Feels From the Inside](#), Eliezer writes about disagreement on definitions. "We know where Pluto is, and where it's going; we know Pluto's shape, and Pluto's mass - but is it a planet?" The question, he says, is meaningless. It's a spandrel from our cognitive algorithm, which works more efficiently if it assigns a separate central variable `is_a_planet` apart from all the actual tests that determine whether something is a planet or not.

Aaron and Zahra seem to be making the same sort of mistake. They have a separate variable `is_a_religion_of_peace` that's sitting there completely separate from all of the things you might normally use to decide whether one group of people is generally more violent than another.

But things get much worse than they do in the Pluto problem. Whether or not Pluto is a planet feels like a factual issue, but turns out to be underdetermined by the facts. Whether or not Islam is a religion of peace feels like a factual issue, but is really a false front for a whole horde of beliefs that have no relationship to the facts at all.

When Zahra says "Islam is a religion of peace," she is very likely saying something along the lines of "I like Islam!" or "I like tolerance!" or "I identify with an in-group who say things like 'Islam is a religion of peace'" or "People who hate Islam are mean!" or even "I don't like Republicans.". She may be covertly pushing policy decisions like "End the war on terror" or "Raise awareness of unfair discrimination against Muslims."

When Aaron says "Islam is not a religion of peace," he is probably saying something like "I don't like Islam," or "I think excessive tolerance is harmful", or "I identify with an in-group who would never say things like 'Islam is a religion of peace'" or even "I don't

like Democrats." He may be covertly pushing policy decisions like "Continue the war on terror" or "Expel radical Muslims from society."

Eliezer's solution to the Pluto problem is to uncover the [disguised query](#) that made you care in the first place. If you want to know whether Pluto is spherical under its own gravity, then without worrying about the planet issue you can simply answer yes. And you're wondering whether to worry about your co-worker Abdullah bombing your office, you can simply answer no. Islam is peaceful enough for your purposes.

But although uncovering the disguised query is a complete answer to the Pluto problem, it's only a partial answer to the religion of peace problem. It's unlikely that someone is going to misuse the definition of Pluto as a planet or an asteroid to completely misunderstand what Pluto is or what it's likely to do (although [it can happen](#)). But the entire point of caring about the "Islam is a religion of peace" issue is so you can misuse it *as much as possible*.

Israel is evil, because it opposes Muslims, and Islam is a religion of peace. The Democrats are tolerating Islam, and Islam is not a religion of peace, so the Democrats must have sold out the country. The War on Terror is racist, because Islam is a religion of peace. We need to ban headscarves in our schools, because Islam is not a religion of peace.

I'm not sure how the chain of causation goes here. It could be (emotional attitude to Islam) -> (Islam [is/isn't] a religion of peace) -> (poorly supported beliefs about Islam). Or it could just be (emotional attitude to Islam) -> (poorly supported beliefs about Islam). But even in the second case, that "Islam [is/isn't] a religion of peace" gives the poorly supported beliefs a dignity that they would not otherwise have, and allows the person who holds them to justify themselves in an argument. Basically, that one phrase holes itself up in your brain and takes pot shots at any train of thought that passes by.

The presence of that extra `is_a_religion_of_peace` variable is not a benign feature of your cognitive process anymore. It's a malevolent mental smuggler transporting prejudices and strong emotions into seemingly reasonable thought processes.

Which brings us back to soft positivism. If we find ourselves debating statements that we refuse to reduce to empirical data<sup>6</sup>, or using statements in ways their reductions don't justify, we need to be extremely careful. I am not positivist enough to say we should never be doing it. But I think it raises one heck of a red flag.

Agree with me? If so, which of the following statements do you think are reducible, and how would you begin reducing them? Which are completely meaningless and need to be scrapped? Which ones raise a red flag but you'd keep them anyway?

1. All men are created equal.
2. The lottery is a waste of hope.
3. Religious people are intolerant.
4. Government is not the solution; government is the problem.
5. George Washington was a better president than James Buchanan.
6. The economy is doing worse today than it was ten years ago.
7. God exists.
8. One impulse from a vernal wood can teach you more of man, of moral evil, and of good than all the sages can.

9. Imagination is more important than knowledge.
10. Rationalists should *win*.

### **Footnotes:**

**1:** More properly the machine running the multiverse, since this would allow counterfactuals to be meaningful. It would also simplify making a statement like "The patient survived because of the medicine", since it would allow quick comparison of worlds where the patient did and didn't receive it. But if the machine is running the multiverse, where's the machine?

**2:** One thing I learned from the comments on Eliezer's post is that this criterion is often very hard to apply in theory. However, it's usually not nearly as hard in practice.

**3:** This sounds like the sort of thing there should already be a name for, but I don't know what it is. Verificationism is too broad, and empiricism is something else. I should point out that I am probably misrepresenting the positivist position here quite badly, and that several dead Austrians are either spinning in their graves or (more likely) thinking that this whole essay is meaningless. I am using "positivist" only as a pointer to a certain *style* of thinking.

**4:** Before this issue dominates the comments thread: yes, I realize that the president having any impact on the economy is highly debatable, that there's not nearly enough data here to make a generalization, et cetera. But my uncle's statement - that Democratic presidents *hurt* the economy, is clearly not supported.

**5:** If your interpretation of anything in the following example offends you, please don't interpret it that way.

**6:** Where morality fits into this deserves a separate post.

# When Truth Isn't Enough

**Continuation of:** [The Power of Positivist Thinking](#)

Consider this statement:

The ultra-rich, who control the majority of our planet's wealth, spend their time at cocktail parties and salons while millions of decent hard-working people starve.

A soft positivist would be quite happy with this proposition. If we define "the ultra-rich" as, say, the richest two percent of people, then a quick look at the economic data shows they do control the majority of our planet's wealth. Checking up on the guest lists for cocktail parties and customer data for salons, we find that these two activities are indeed disproportionately enjoyed by the rich, so that part of the statement also seems true enough. And as anyone who's been to India or Africa knows, millions of decent hard-working people do starve, and there's no particular reason to think this isn't happening at the same time as some of these rich people attend their cocktail parties. The positivist scribbles some quick calculations on the back of a napkin and certifies the statement as **TRUE**. She hands it the Official Positivist Seal of Approval and moves on to her next task.

But the truth isn't always enough. Whoever's making this statement has a much deeper agenda than a simple observation on the distribution of wealth and preferred recreational activities of the upper class, one that the reduction doesn't capture.

Philosophers like to speak of the denotation and the connotation of a word. Denotations (not to be confused with [denettations](#), which are much more fun) are simple and reducible. To capture the denotation of "old", we might reduce it to something testable like "over 65". Is Methusaleh old? He's over 65, so yes, he is. End of story.

Connotations<sup>0</sup> are whatever's left of a word when you subtract the denotation. Is Methusaleh old? How dare you use that word! He's a "senior citizen!" He's "elderly!" He's "in his golden years." Each of these may share the same denotation as "old", but the connotation is quite different.

There is, oddly enough, a children's game about connotations and denotations<sup>1</sup>. It goes something like this:

I am intelligent. You are clever. He's an egghead.  
I am proud. You are arrogant. He's full of himself.  
I have perseverance. You are stubborn. He is pig-headed.  
I am patriotic. You're a nationalist. He is jingoistic.

Politicians like this game too. Their version goes:

I care about the poor. You are pro-welfare. He's a bleeding-heart.  
I'll protect national security. You'll expand the military. He's a warmonger.  
I'll slash red tape. You'll decrease bureaucracy. He'll destroy safeguards.  
I am eloquent. You're a good speaker. He's a demagogue.

I support free health care. You support national health care. He supports socialized health care.

All three statements in a sentence have the same denotation, but very different connotations. The Connotation Game would probably be good for after-hours parties at the Rationality Dojo<sup>2</sup>, playing on and on until all three statements in a trio have mentally collapsed together.

Let's return to our original statement: "The ultra-rich, who control the majority of our planet's wealth, spend their time at cocktail parties and salons while millions of decent hard-working people starve." The denotation is a certain (true) statement about distribution of wealth and social activities of the rich. The connotation is hard to say exactly, but it's something about how the rich are evil and capitalism is unjust.

There is a serious risk here, and that is to start using this statement to build your belief system. Yesterday, I suggested that saying "Islam is a religion of peace" is meaningless but affects you anyway. Place an overly large amount of importance on the "ultra-rich" statement, and it can play backup to any other communist beliefs you hear, even though it's trivially true and everyone from Milton Friedman on down agrees with it. The associated Defense Against The Dark Arts technique is [to think like a positivist](#), so that this statement and its reduced version sound equivalent<sup>3</sup>.

...which works fine, until you get in an argument. Most capitalists I hear encounter this statement will flounder around a bit. Maybe they'll try to disprove it by [saying something very questionable](#), like "If people in India are starving, then they're just not working hard enough!" or "All rich people deserve their wealth!"<sup>4</sup>

Let us take a moment to feel some sympathy for them. The statement sounds like a devastating blow against capitalism, but the capitalists cannot shoot it down because it's technically correct. They are forced to either resort to peddling falsehoods of the type described above, or to sink to the same level with replies like "That sounds like the sort of thing *Stalin* would say!" - which is, of course, denotatively true.

What would I do in their position? I would stand tall and say "Your statement is technically true, but I disagree with the connotations. If you state them explicitly, I will explain why I think they are wrong."

YSITTBIDWTCIYSTEIWEWITTAW is a little long for an acronym, but ADBOC for "Agree Denotationally But Object Connotationally" could work. *[EDIT: Changed acronym to [better suggestion by badger](#)]*

## Footnotes

**0:** Anatoly Vorobey [says in the comments](#) that I'm using the word connotation too broadly. He suggests "subtext".

**1:** I feel like I might have seen this game on Overcoming Bias before, but I can't find it there. If I did, apologies to the original poster.

**2:** Comment with any other good ones you know.

**3:** Playing the Connotation Game a lot might also give you partial immunity to this.



**4:** This is a great example of a hotly-debated statement that is desperately in need of reduction.