

Rationality without Absolutes: How to Think like a Bayesian*

MICHAEL G. TITELBAUM, Philosophy, University of Wisconsin, Madison.

[Michael G Titelbaum](#) is Vilas Distinguished Achievement Professor of Philosophy at the University of Wisconsin-Madison. He is the author of [Quitting Certainties](#) (2013) and [Fundamentals of Bayesian Epistemology, Volume 1 and Volume 2](#) (2022).



* Essay first appeared in [Aeon Magazine](#), 08 January 2024. Reproduced with permission.

You're often asked what you believe. Do you believe in God? Do you believe in global warming? Do you believe in life after love? And you're often told that your beliefs are central to who you are, and what you should do: 'Do what you believe is right.'

These belief-questions demand all-or-nothing answers. But much of life is more complicated than that. You might not believe in God, but also might not be willing to rule out the existence of a deity. That's what agnosticism is for.

For many important questions, even three options aren't enough. Right now, I'm trying to figure out what kinds of colleges my family will be able to afford for my children. My kids' options will depend on lots of variables: what kinds of schools will they be able to get into? What kinds of schools might be a good fit for them? If we invest our money in various ways, what kinds of return will it earn over the next two, five, or 10 years?

Suppose someone tried to help me solve this problem by saying: 'Look, it's really simple. Just tell me, do you believe your oldest daughter will get into the local state school, or do you believe that she won't?' I wouldn't know what to say to that question. I don't believe that she *will* get into the school, but I also don't believe that she *won't*. I'm perhaps slightly more confident than 50-50 that she will, but nowhere near certain.

One of the most important conceptual developments of the past few decades is the realisation that *belief comes in degrees*. We don't just believe something or not: much of our thinking, and decision-making, is driven by varying levels of confidence. These confidence levels can be measured as probabilities, on a scale from zero to 100 per cent. When I invest the money I've saved for my children's education, it's an oversimplification to focus on questions like: 'Do I believe that stocks will outperform bonds over the next decade, or not?' I can't possibly know that. But I can try to assign educated probability estimates to each of those possible outcomes and balance my portfolio in light of those estimates.

We know from many years of studies that reasoning with probabilities is hard. Most of us are raised to reason in all-or-nothing terms. We're quite capable of *expressing* intermediate degrees of confidence about events (quick: how confident are you that a Democrat will win the next presidential election?), but we're very bad at *reasoning* with those probabilities. Over and over, studies have revealed systematic errors in ordinary people's probabilistic thinking.

Luckily, there once lived a guy named the Reverend [Thomas Bayes](#). His work on probability mathematics in the 18th century inspired a movement we now call Bayesian statistics. You may have heard 'Bayesian' talk thrown around in conversation, or mentioned in news articles. At its heart, Bayesianism is a toolkit for reasoning with probabilities. It tells you how to measure levels of confidence numerically, how to test those levels to see if they make sense, and then how to manage them over time.

That last part is important because, for any given claim, you're more confident in it at some times than you are at others. Once my oldest daughter takes a bunch of standardised tests, I'll have new evidence about her college prospects, and will adjust my levels of confidence accordingly. Bayesianism provides a recipe for doing that.

In this Guide, I'll provide five basic Bayesian ideas to improve your reasoning with probabilities. They're just a start – if you really want to delve into the details and grapple with the mathematics, I'll give you links and book recommendations at the end. I can't promise to make you a perfect probabilistic reasoner. (Heck, I teach this stuff for a living, and I still make mistakes.) But hopefully this will give you a start, helping you better apportion your opinions to your evidence, and make better decisions in the face of uncertainty.

Embrace the margins

The first step to Bayesianism is to stop thinking in all-or-nothing terms. Bayesians want to move past the dichotomy of you-either-believe-it-or-you-don't, to start thinking of belief as something that comes in degrees. Those degrees can be measured on a zero to 100 per cent scale. If you're certain an event will occur, that's 100 per cent confidence. If you're certain it won't occur, that's 0 per cent.

But again, Bayesians counsel against going to extremes. There are very few situations in which it makes sense to be *certain* that something will happen, or that it won't. In his book [Making Decisions](#) (1971), the Bayesian Dennis Lindley approvingly cited Oliver Cromwell's dictum to always 'think it possible that you may be mistaken.' Unless an event is strictly *impossible*, you shouldn't be certain that it won't occur.

All right, fine then. Maybe we shouldn't assign anything that's strictly speaking possible a confidence of zero. But we've all heard someone describe a possibility as 'one in a million'. If something's that improbable, it's pretty much not going to happen, right? So, one in a million might as well be zero? The very same Dennis Lindley also said he was fine assigning a confidence of one in a million that the Moon is made of green cheese.

A common mistake when reasoning with probabilities is to think that a fraction of a percentage point – especially near such extreme values as 0 per cent or 100 per cent – really doesn't matter. Any parent who's been fortunate enough to get high-quality modern-day prenatal care will have seen genetic tests reporting how likely their growing fetus is to develop certain kinds of ailments and birth defects. I remember looking at probabilities like 0.0004 per cent and 0.019 per cent with my pregnant wife, and wondering what we should be worried about and what we could write off. Such small probability differences are difficult to grasp intuitively. But a condition with a probability of 0.019 per cent is almost 50 times as likely to occur as one with a probability of 0.0004 per cent.

It's tempting to see a probability value like 0.0001 per cent – one in a million – and assume the difference between that and 0 per cent is little more than a rounding error. But an event with 0 per cent probability literally can't happen, while events with a probability of 0.0001 per cent happen all the time. If you have a couple of minutes and some loose change, go flip a coin 20 times. (We'll wait.) Whatever sequence of heads and tails you wound up observing, that specific sequence had a less than one-in-a-million chance of occurring.

To better assess the significance of the almost impossible and the almost certain, Bayesians sometimes switch from measuring probabilities on a percentage scale to measuring them with odds. If I bought you enough tickets to have a 0.001 per cent chance at winning the lottery and bought your friend enough tickets to give him a 0.1 per cent chance, you might wonder how offended you should be. Putting those values in odds form, we see that I've given your friend a 1 in 1,000 shot and you only 1 in 100,000 shot! Expressing the probabilities in odds form makes it clear that your friend has 100 tickets for every 1 of yours and clarifies that these two probabilities – while admittedly both close to zero – are nevertheless importantly different.

Evidence supports what makes it probable

What did the Rev Bayes do to get a whole statistical movement named after him? Prior to Bayes, much probability theory concerned problems of 'direct inference'. This is the kind of probability problem you were asked to solve many times in school. You're told that two fair, six-sided dice are rolled, and are asked to calculate the probability that their sum will be eight. Put a bit more abstractly: you're given a hypothesis about some probabilistic process in the world and asked to compute the probability that it will generate a particular kind of evidence.

Bayes was interested in the opposite: so-called 'inverse inference'. Suppose you observe some evidence and want to infer back to a hypothesis about what kind of process in the world might have generated that evidence. In [*The Theory of Probability*](#) (1935), Hans Reichenbach listed many occasions on which we engage in reasoning with this structure:

The physician's inferences, leading from the observed symptoms to the diagnosis of a specified disease, are of this type; so are the inferences of the historian determining the

historical events that must be assumed for the explanation of recorded observations; and, likewise, the inferences of the detective concluding criminal actions from inconspicuous observable data.

Bayes's most important contribution to inverse inference wasn't recognised during his lifetime. After the reverend died in 1761, a Welsh minister named [Richard Price](#) published a theorem he had found in Bayes's notes. This theorem was later independently rediscovered by Pierre-Simon Laplace, who was responsible for much of its early popularisation.

Price, Laplace and others promoted [Bayes's Theorem](#) as a rule for adjusting one's confidence in a hypothesis after discovering some new piece of evidence. Modern Bayesians are called 'Bayesians' because of their adherence to Bayes's Rule (or Theorem). According to Bayes's Rule, your updated confidence in the hypothesis should be calculated from two factors: what your confidences looked like *before* you got the evidence (about which more later), and how strongly the evidence supports the hypothesis.

Here it pays to remember Bayesians' aversion to absolutes. While it makes for good drama when a character learns a single piece of information that changes their whole worldview, most of life isn't like that. Each new piece of information we gain changes only some of our opinions and changes them incrementally – making us slightly more confident or slightly less confident that particular events will occur. This is because evidential support also comes in degrees: a piece of evidence might support some hypotheses weakly and others strongly; or one piece of evidence might support a particular hypothesis better than another.

To gauge how strongly evidence supports some hypothesis, ask how likely that hypothesis makes the evidence. Suppose you get home late from work one night and walk in to find all the lights on in your home. You wonder who else is home – your husband? Your son? Well, your husband is constantly griping about the power bills, and walks around the house turning lights off all the time. But your teenage son barely notices his surroundings and exits a room without a thought to how he's left it. The evidence you've found is very likely if your son is in the house, and much less likely if your husband is home. So, the evidence supports your son's presence strongly, and your husband's presence little or not at all.

Bayes's Rule says that, once you gauge how much your new evidence supports various hypotheses, you should shift your confidence towards hypotheses that are better supported. However confident you were before you walked in the door that your husband or son was home, what you find inside should increase your confidence that your son is there, and decrease your confidence that your husband is. How much increase and decrease are warranted? That's all sorted out by the specific mathematics of Bayes's Rule. I'm trying to keep it light here and avoid equations, but the sources in the final section can fill in the details.

Attend to *all* your evidence

A consistent theme of Bayesian thinking is that working in shades of confidence can get much more complex and subtle than thinking in absolutes. One of the nice features of conclusive, slam-dunk evidence is that it can't be overridden by anything. If a mathematician proves some theorem, then nothing learned subsequently can ever undo that proof, or give us reason not to believe its conclusion.

Bayesianism aims to understand incremental evidence, coming to terms with the kinds of less-than-conclusive information we face every day. One crucial feature of such evidence is that it can *always* be overridden. This is the lifeblood of twisty mystery novels: an eyewitness said the killer held the gun in his *left* hand – but it turns out she was looking in a mirror – but the autopsy reveals the victim was poisoned *before* he was shot...

Because the significance of evidence depends so much on context, and because potential defeaters might always be lurking, it's important not to become complacent with what one knows and to keep an open mind for relevant new information. But it's also important to think thoroughly and carefully about the information one already has. [Rudolf Carnap](#) proposed the Principle of Total Evidence, which requires your beliefs about a question to incorporate and reflect *all* the evidence you possess relevant to that question.

Here's a kind of relevant evidence we often overlook: besides having information about some topic, we often know something about how we got that information. Now, that's not always true: I know that Abraham Lincoln was born in a log cabin, but I have no idea where I learned that tidbit. But often – and especially in today's uncertain media environment – it pays to keep track of one's sources, and to evaluate whether the information you've received might have been selected for you in a biased way.

Sir [Arthur Eddington](#) gave an example in which you draw a large group of fish from a lake, and all of them are longer than six inches. Normally, this would be strong evidence that all the fish in the lake are at least that long. But if you know that you drew the fish using a net with six-inch holes, then you can't draw what would otherwise be the reasonable conclusion from your sample.

Paying attention to how the evidence was selected can have important real-life consequences. In [How Not to Be Wrong](#) (2014), Jordan Ellenberg recounts a story from the Second World War: the US military showed the statistician Abraham Wald data indicating that planes returning from dogfights had more bullet holes in the fuselage than in the engine. The military was considering shifting armour from the engine to the fuselage, to better protect their pilots. Wald recommended exactly the opposite, on the grounds that it was the *returning* planes that had holes in the fuselage; those not returning had holes to their engines, so that's where the additional armour should go.

Don't forget your prior opinions

You think carefully about the evidence you've just received. You're careful to take it *all* into account, to consider context, and to remember where it came from. With all this in mind, you find the hypothesis that renders that evidence most probable, the hypothesis most strongly supported by that evidence. That's the hypothesis you should now be most confident in, right?

Wrong.

Bayes's Rule says to respond to new evidence by *increasing* your confidence in the hypothesis that makes that evidence most probable. But where you land after an increase depends on where your confidence was before that evidence came in.

Adapting an example from the reasoning champion Julia Galef, suppose you're crossing a college campus and stop a random undergraduate to ask for directions. This undergrad has a distracted, far-off look in their eye; wears clothes that one would never think of bringing near an iron; and seems slightly surprised to even be awake at this hour of the day. Should you be more confident that your interlocutor is a philosophy or a business major?

Easy answer: this look is much more typical of a philosophy major than a business major, so you should be more confident you're dealing with the former. At a first pass, that answer seems backed up by the Bayesian thinking I've described. Just to pick some numbers (and be a bit unfair to philosophers), let's suppose a third of all philosophy majors meet this description, but only one in 20 business majors does (the quants, perhaps?) On the hypothesis that the person you randomly stopped for directions is a philosophy major, the probability of your evidence is one-third. On the hypothesis that you stopped a business major, the probability is one-20th. So, your evidence about this student from their appearance more strongly supports the notion that they study philosophy.

But now consider the following: on my campus, there are currently just shy of 250 undergraduate philosophy majors and roughly 3,600 business majors. If the fractions in the previous paragraph are correct, we should expect there to be about 80 philosophy students on campus disconnected from their surroundings, and about 180 business majors. So, if you select a random undergrad, you're still at least twice as likely to get a distracted business major as a distracted philosopher.

The key here is to remember that, before you appraised this student's appearance, the odds were much, much greater that they were into business than philosophy. The evidence you gain from interacting with them should *increase* your confidence that they're a philosopher, but increasing a small number can still leave it quite small!

Bayes's Rule demands that your updated confidence in a hypothesis after learning some evidence combines *two* factors: your prior confidence in the hypothesis, and how strongly it's supported by the new evidence. Forgetting the former, and attending only to the latter, is known as the [Base Rate Fallacy](#). Unfortunately, this fallacy is committed frequently by professionals, even those working with life-altering data.

Suppose a new medical test has been developed for a rare disease – only one in 1,000 people has this disease. The test is pretty accurate: someone with the disease will test positive 90 per cent of the time, while someone without the disease will test positive only 10 per cent of the time. You randomly select an individual, apply the test, and get a positive result. How confident should you be that they have the disease?

Most people – including trained medical professionals! – say you should be 80 per cent or 90 per cent confident that the individual has the disease. The correct answer, according to Bayes's Rule, is under 1 per cent. What's going on is that most respondents are so overwhelmed by the accuracy of the test (the strength of the evidence it produces) that they neglect how rare this disease is in the population.

But let's do some quick calculations: suppose you applied this test to 10,000 randomly selected individuals. Around 10 of them would have the disease, so nine of them would get a positive test result. On the other hand, around 9,990 of the individuals you selected wouldn't have the disease. Since the test gives healthy individuals a positive result 10 per cent of the

time, these 9,990 healthy individuals would yield around 999 false positive tests. So having tested 10,000 people, you'd get a total of 1,008 positive results, of which only nine (just under 1 per cent) would be people who actually had the disease.

Again, when dealing with cases of extreme probabilities, it can help to think about the odds. A piece of evidence that strongly supports a hypothesis (like the reliable medical test just described) might multiply the odds of that hypothesis by a factor of 10, or even 100. But if the odds start small enough, multiplying them by 10 will take you from one chance in 1,000 to one in 100.

Subgroups don't always reflect the whole

Bayesians work a lot with conditional probabilities. Conditional probability arises when you consider how common some trait is among a subgroup of the population, instead of considering the population as a whole. If you pick a random American, they're very unlikely to enjoy pizza made with an unleavened crust, topped with Provel cheese, and cut into squares. But conditional on the assumption that they grew up in St Louis, the probability that they'll enjoy such a monstrosity is much higher.

Conditional probabilities can behave quite counterintuitively. Simple principles that one would think should be obvious can fail in spectacular fashion. The clearest example of this is [Simpson's Paradox](#).

Hopefully all of us have learned in our lives not to draw broad generalisations from a single example, or to assume that a small group is representative of the whole. A foreigner who judged American pizza preferences by visiting only St Louis would be seriously misled. By carelessness or sheer bad luck, we can stumble into a subpopulation that is unlike the others, and so bears traits that aren't reflected by the population in general.

But Simpson's Paradox demonstrates something much weirder than that: sometimes *every* subpopulation of a group has a particular trait, but that trait still isn't displayed by the group as a whole.

In the 2016-17 NBA season, James Harden (then of the Houston Rockets) made a higher percentage of his two-point shot attempts than DeMar DeRozan (of the Toronto Raptors) made of *his* two-point shots. Harden also sank a higher percent of his three-point attempts than DeRozan. Yet DeRozan's overall field-goal percentage – the percent of two-pointers and three-pointers combined that he managed to sink – was higher than Harden's. Harden did better on both two-pointers and three-pointers, and those are the only kinds of shots that factor into the field-goal percentage, yet DeRozan was better overall. How is that possible?

Pro hoops aficionados will know that, for any player, two-point shots are easier to hit than three-pointers, yet Harden stubbornly insists on making things difficult for himself. In the 2016-17 season, he attempted almost the same number of each kind of shot (777 three-pointers versus 756 two-pointers), while DeRozan attempted more than 10 times as many two-pointers as three-pointers. Even though Harden was better at each kind of shot, DeRozan made the strategic decision to take high-percentage shots much more often than low-percentage ones. So, he succeeded at an overall higher rate.

The same phenomenon appeared when graduate departments at the University of California, Berkeley were investigated for gender bias in the 1970s. In 1973, 44 per cent of male applicants were admitted to Berkeley's graduate school, while only 35 per cent of female applicants succeeded. Yet a statistical study found that individual departments (which actually made the admissions decisions) were letting in men and women at roughly equal rates, or even admitting women more often. The trouble was that some departments were much more difficult than others to get into (for all applicants!), and women were applying disproportionately to more selective fields.

Of course, that doesn't eliminate all possibilities of bias; a study found that women were applying to more crowded fields because they weren't given the undergraduate mathematical background to study subjects that were better-funded (and therefore could admit more students). But the broader point about conditional probabilities stands: You can't assume that an overall population reflects trends in its subpopulations, even if those trends occur in *all* the subpopulations. You also have to consider the distribution of traits *across* subpopulations.

Why it matters

Bishop Joseph Butler said: 'Probability is the very guide of life.' Rev Bayes taught us to use that guide and update it over time as our lives change and we learn new things.

Bayes's Rule is an equation; if you want the numerical details, you can find them in the sources at the end of the Guide. But the basic recipe for updating your confidences is: Start with your prior opinions. Consider your new evidence – *everything* you just learned, including what you know about how you learned it. Of the hypotheses you entertain, determine which make that evidence more probable. Then shift your confidence towards those.

You might ask: where do the prior opinions come from? If you're a Bayesian, the opinions you take into a particular investigation will have been influenced by evidence you gathered in the past. You don't just apply Bayes's Rule once. Every time you gain new information about a subject, you update your opinions on that subject, with those newly updated opinions supplying the priors for your next update in the future. Your ongoing, ever-evolving picture of the world is like [Otto Neurath](#)'s image of the boat:

'[W]e are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom...'

No two people ever have the same course of evidence, and no two people ever have the same sequence of opinions over their lives. We should keep these divergent paths in mind when we encounter different views. But we should also remember one beautiful piece of Bayesian mathematics: If we apply Bayes's Rule every time we update our opinions, then, no matter where our opinions begin, there's a high probability that gathering more and more evidence will move them ever closer towards the truth. If we keep learning, and keep updating, then Bayes's guide will lead us to our destination.

Recapitulation: Thinking Like a Bayesian

1. **Embrace the margins.** It's rarely rational to be certain of anything. Don't confuse the improbable with the impossible. When thinking about extremely rare events, try thinking in odds instead of percentages.
2. **Evidence supports what makes it probable.** Evidence supports the hypotheses that make the evidence likely. Increase your confidence in whichever hypothesis makes the evidence you're seeing most probable.
3. **Attend to *all* your evidence.** Consider all the evidence you possess that might be relevant to a hypothesis. Be sure to take into account *how* you learned what you learned.
4. **Don't forget your prior opinions.** Your confidence after learning some evidence should depend both on what that evidence supports and on how you saw things before it came in. If a hypothesis is improbable enough, strong evidence in its favour can still leave it unlikely.
5. **Subgroups don't always reflect the whole.** Even if a trend obtains in *every* subpopulation, it might not hold true for the entire population. Consider how traits are distributed *across* subgroups as well.

Links & books

Many of the lessons and examples in this piece were taken from my book [Fundamentals of Bayesian Epistemology](#) (2022). That text takes you through all the mathematical details, teaches you to apply Bayesianism to decision theory and the theory of evidential support, and contrasts Bayesianism with rival statistical schools.

A less detailed treatment, written at a more introductory level, is Jonathan Weisberg's online, open-source [Odds & Ends](#). Along similar lines is Darren Bradley's [A Critical Introduction to Formal Epistemology](#) (2015).

The online Stanford Encyclopedia of Philosophy has an excellent article called [Bayesian Epistemology](#) by Hanti Lin, that will give you loads of information (mathematical, philosophical, argumentative) about the subject without your having to read an entire book.

Among the many philosophical videos on her YouTube channel [Measure of Doubt](#), [Julia Galef](#) has a number of good ones on Bayesian thinking. She's especially skilled at illustrating the relevant numerical manipulations with clear diagrams. Galef also hosts the podcast *Rationally Speaking*. Some of its early episodes, with co-host Massimo Pigliucci, answer questions about Bayesianism.