# Bayesianism

Author: Thomas Metcalf
Categories: Epistemology, Philosophy of Science, Logic and Reasoning
Word Count: 1000

Editor's note: for many readers, this introduction to Bayesianism would be more profitably read after first reading Introduction to the Probability Calculus by Thomas Metcalf.

Bayesianism says that degrees of belief or justification can be represented by probabilities, and that we can assess the rationality of degrees of belief— of *credences*—by seeing whether they follow a certain set of rules.[1] This essay is an introduction to Bayesianism.[2]
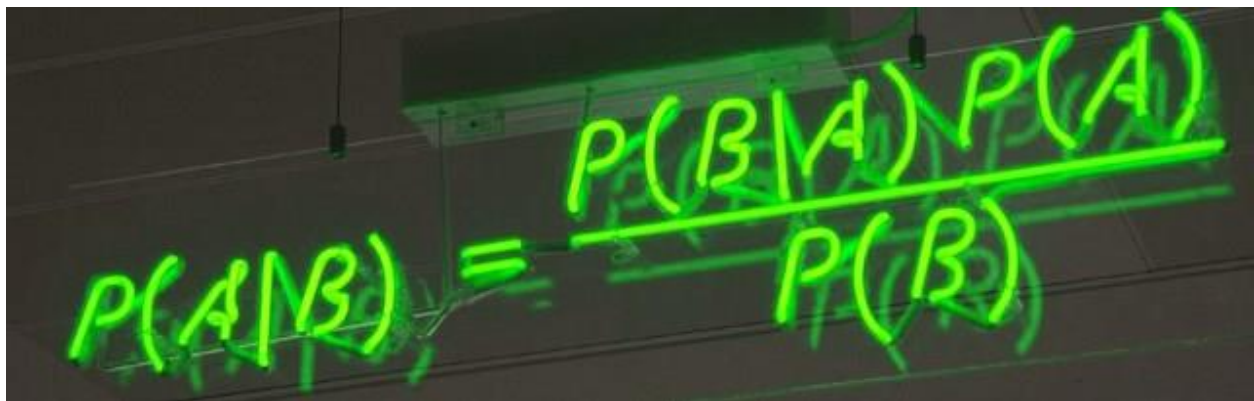
An image of a neon-style sign of Bayes's Theorem.

1. An Example

Suppose you draw a card randomly out of a standard, shuffled deck and lay it facedown on the table without looking at it. What's the probability that this Facedown Card is an ace?

There are four aces among the fifty-two cards in a standard deck, and each card is equally likely to have been drawn, so the probability is 4/52 or about 7.7%.

This 7.7% is the *prior* probability that the Facedown Card is an ace: prior to drawing any more cards.

Now, leaving the Facedown Card unviewed, you draw ten *more* cards out of the deck and look at them; *none* is an ace. So there are forty-two cards remaining unviewed, including all four aces: forty-one cards still in the deck, plus the Facedown Card.

Now what's the probability that the Facedown Card is an ace? There are now forty-two *unviewed* cards; ten *viewed* cards (none of which are aces); and four aces still among the forty-two *unviewed* cards. So the probability is 4/42 ≈ 9.5%.[3]

This "9.5%" is the *posterior* probability that the Facedown Card is an ace: posterior to (i.e., *after*) gathering the evidence of ten non-aces.

In this process, we began with a probability of 7.7% that the Facedown Card was an ace. Then you gathered new evidence, and the probability became 9.5%.

2. Probabilities and Credences

Bayesians talk about *personal* probabilities: degrees of belief, credence, confidence, or justification held by a particular person.[4] If I'm certain of some hypothesis *H*, then I have 100% credence in *H*. If I'm certain it's false, then I have 0% credence.

In the story about cards, arguably, it would be *rational* to begin with 7.7% credence that the Facedown Card was an ace.[5] Then, after drawing ten non-aces, it would be rational to *update* your credence to 9.5% that the Facedown Card was an ace.[6]

3. Bayes's Rule

Your credence that the Facedown Card would be an ace should start at a certain point (7.7%) and then change based on new evidence, as we saw above. How exactly should it change?

The following answer is a defining tenet of Bayesianism:

Bayes's Rule: If you acquire some evidence E_1, then your new credence $P_{NEW}(H)$ in hypothesis H (after learning E_1) should become equal to the old credence $P_{OLD}(H|E_1)$, i.e., the probability of H given that E_1 is true.[7]

Your credences were previously captured by probability function P_{OLD}, according to which $P_{OLD}(H) = 7.7\%$, and $P_{OLD}(H|E_1) = 9.5\%$. When you learned E_1, you updated your beliefs to a new probability function P_{NEW} according to which $P_{NEW}(H) = P_{OLD}(H|E_1) = 9.5\%$.

Then, if we acquire further evidence (e.g., by drawing even more cards), we treat that $P_{OLD}(H|E_1) = P_{NEW}(H)$ as our $P(H)$ for *further* calculations based on some even-*newer* evidence E_2. Term $P_{OLD}(H|E_2)$ will be the new $P(H)$ for when we encounter E_3, and so on.

By this process, rational cognizers repeatedly update their credence in some hypothesis.[8]

4. Bayes's Theorem

It's nice to have Bayes's Rule, but employing it requires answering a big question: How do we calculate $P(H|E)$? That is, how do we know what credence to put in our hypothesis after encountering our evidence?

Fortunately, we have Bayes's Theorem. There are several useful versions of the theorem,[9],[10] but we'll discuss this one:[11]

Bayes's Theorem:[12] $P(H|E) = P(E|H) \times P(H) / P(E)$, where $P(E) \neq 0$.

The theorem tells us what credence to put in hypothesis *H* after we acquire evidence *E*. The credence, as you can see, is based on the *prior* (to encountering *E*) probabilities of *H* and *E*, and based on the *likelihood*[13] P(E|H): how probable it is that *E* would be the case given that *H* is the case.

The theorem is intuitive:

- The more surprising the evidence is—that is, the lower the *prior* probability P(E)—the more the evidence supports the hypothesis: the higher P(H|E).

- The more probable the hypothesis already was—that is, the higher the *prior* P(H)—the more probable it should be *after* encountering *E* too: the higher P(H|E).

- And the better the hypothesis predicts the evidence—that is, the higher P(E|H), which is the *likelihood* of *E* given *H*—the more probable *H* should be after acquiring *E*: the higher P(H|E).

5. Applying Bayes's Theorem to the Example of Cards

Consider the hypothesis "*A*" that the Facedown Card is an ace, and the evidence "*T*" that you drew out ten non-ace cards from the deck. Then we want to know P(A|T). Consider:[14]

- We already calculated P(A): that's the *prior* probability that the Facedown Card was an ace, before drawing those ten cards out: about 7.7%.

- If the Facedown Card *is* an ace—that is, given our hypothesis *A*—then there are 51 cards remaining in the deck before drawing more, and *three* of them are aces. The *likelihood* P(T|A) that you would draw out *ten* non-aces (without replacement) from a 51-card set containing *three* aces can be calculated: it's about 51.2%.[15]

- The prior probability P(T) that you would draw out ten non-aces from fifty-one cards is about 41.3%.[16]

So calculate:

P(A|T) = P(T|A) × P(T) / P(A) ≈ 0.512 × 0.413 / 0.077 ≈ 0.095 = 9.5%.

And that's what we saw in Section 1 above. Following the rules leads you to end up at the correct final credence.

In the story about cards, there was an easy calculation available for P(A|T). Yet in real life, we often have access to priors and a likelihood, but it's not so easy to calculate the posterior.[17] In those cases, Bayes's Theorem is extremely useful.

6. Conclusion

Bayesianism has enormous popularity and many useful applications across a wide variety of disciplines.[18] However, there are still many points of dispute among Bayesians, and challenges for the overall theory.[19]

Notes

[1] The main rules are the Kolmogorov axioms and Bayes's Rule; see below. Strictly speaking, probabilism (the view that credences should obey, or do obey, the Kolmogorov axioms of probability) is a tenet of epistemic Bayesianism, but not the only tenet. See Joyce (2005, p. 153). For more on these rules and axioms, see Introduction to the Probability Calculus by Thomas Metcalf. Bayesianism also says that probabilities in science, especially in statistics, should be understood as personal or belief-like (not objective, physical) probabilities; see Weisberg (n.d., ch. 15) and Interpretations of Probability by Thomas Metcalf.

[2] Bayesianism is named after the philosopher and statistician the Rev. Thomas Bayes (c. 1701–1761). His presentation of the theorem appeared in 1763 (Bayes & Price, 1763).

For recent introductory or overview works on Bayesianism and Bayesian epistemology, see e.g. Hacking (2001), Joyce (2005), Howson & Urbach (2006), Steinhart (2009, ch. 5), Weisberg (2011), Carr (2013), Huber (2019), Huber (n.d.), Schupbach (2022), Talbott (2022), and Weisberg (n.d.).

[3] In case this is puzzling, think about it from a differently ordered process. Suppose I pull out ten non-aces from a deck; shuffle the remaining forty-two-card deck; and then draw out an additional card and lay it facedown. In this case, like in the original process, that facedown card is (4/42)-probable to be an ace.

[4] For more about types of probability, see Interpretations of Probability by Thomas Metcalf. On the probability in question, see also Hacking (2001, chs. 13–15), Steinhart (2009, sect. 7.2), Huber (2019, chs. 7–8), Shupbach (2022, p. 1), Talbott (2022, sect. 2), and Weisberg (n.d., pt. III). One way to measure someone's degree of confidence is to see which bets they consider to be fair; see, e.g., Weisberg (n.d., sect. 16.1).

[5] This seems to follow from what Lewis (1980) calls the "Principal Principle": If you know that the probability of some outcome is p, then you should be about p-confident that that outcome occurred (cf. Schupbach, 2022, p. 55). One popular way to argue for the rationality of following the rules is to argue that if you don't follow the rules, you could be "tricked" into gambling away all your money. These arguments are sometimes called "Dutch book arguments." See Dutch Book Arguments by Daniel Peterson, along with Talbott (2022, sect. 3) and Vineberg (2022).

[6] Some Bayesians are "subjective Bayesians" and some are "objective Bayesians" (Weisberg, 2011, sect. 3). For subjective Bayesians, you merely have to have credences that obey the Kolmogorov axioms of probability (see Introduction to the Probability Calculus by Thomas Metcalf). For objective Bayesians, there are further rational constraints on one's *prior* probabilities: how confident you should be in some hypothesis H before gathering further evidence (Howson & Urbach, 2006; Talbott, 2022, sect. 4.2.F). Subjective Bayesians argue that in the

long run, roughly speaking (Schupbach, 2022, sect. 2.2.3), if you continue to conditionalize correctly on your evidence, your credence will get closer and closer to objectively accurate. As Joyce (2005, pp. 157–158) observes, most real-life Bayesians will admit some subjective judgments and require others to be objective; Weisberg (2011, sect. 3) concurs that subjectivity and objectivity are on a continuum. For an interesting criticism of subjectivism, namely that it leads to (or is) a form of skepticism, see Huemer (2017, sect. 3).

[7] This rule is sometimes called "conditionalization." See Steinhart (2009, p. 129) and Schupbach (2022, p. 37). See also Talbott (2022), who calls this a "simple principle of conditionalization" (sect. 2). One way to calculate a conditional probability is the following equation: Definition of Conditional Probability: $P(A|B) = P(A\&B) / P(B)$. For more on conditional probability, see <u>Introduction to the Probability Calculus</u> by Thomas Metcalf.

[8] It might be worth specifying that we're imagining that a perfectly rational being would use this process. In real life, there are many basically-rational beings who have never heard of Bayes's Rule, and may not set their posterior probabilities to exactly the correct number, down to the hundredth of a percentage point. Still, Bayesians believe, the closer you are to following Bayes's Rule exactly, the more accurate your posterior credences will be.

[9] Another useful version: Bayes's Theorem (for two mutually-exclusive, jointly exhaustive hypotheses H_1 and H_2): $P(H_1|E) = P(E|H_1) \times P(H_1) / [P(E|H_1) \times P(H_1) + P(E|H_2) \times P(H_2)]$. Notice that we could replace "H_2" with "not H_1," which makes the version especially useful. (You might notice that this version follows from the Law of Total Probability: $P(E) = P(H_1\&E) + P(H_2\&E) = P(E|H_1) \times P(H_1) + P(E|H_2) \times P(H_2)$. See <u>Introduction to the Probability Calculus</u> by Thomas Metcalf.)

[10] Another useful version: Bayes's Theorem ("Odds Form") (for two mutually-exclusive hypotheses): $P(H_1|E) / P(H_2|E) = [P(H_1) \times P(E|H_1)] / [P(H_2) \times P(E|H_2)]$, where $P(H_2)$, $P(H_2|E)$, and $P(E|H_2) \neq 0$. You might notice that this follows from the standard or "probability" form of Bayes's Theorem plus the fact that the two

hypotheses are mutually exclusive (Downey, 2012, sect. 5.2). If $P(H_1|E) = P(E|H_1) \times P(H_1) / P(E)$, then we can divide both sides by $P(H_2|E)$, yielding $P(H_1|E) / P(H_2|E) = [P(E|H_1) \times P(H_1)] / [P(E) \times P(H_2|E)]$. Then, by Bayes's Theorem (standard or probability form), $P(H_2|E) \times P(E) = P(E|H_2) \times P(H_2)$. In turn, we can replace "$P(E) \times P(H_2|E)$" with "$P(E|H_2) \times P(H_2)$," and we get the aforementioned Odds Form of Bayes's Theorem. We sometimes speak of the "odds" of H_1 given E in this case, which is simply $P(H_1|E) / P(H_2|E)$, and the "odds" of H_1 are $P(H_1) / P(H_2)$. In turn, the odds $O(H_1|E)$ of H_1 versus H_2 given E will be equal to the odds $O(H_1)$ of H_1 times the ratio of the likelihoods $P(E|H_1)$ to $P(E|H_2)$. Thus, finally, we sometimes write, "$O(H_1|E) = O(H_1) \times [P(E|H_1) / P(E|H_2)]$."

[11] This version is most-commonly discussed, partly because it is very simple to state, and partly because it zeroes in on exactly what we often want to know: How confident should I be in my hypothesis H now that I've encountered evidence E?

[12] You might notice that this theorem follows from the rule for the probability of a conjunction; see see Introduction to the Probability Calculus by Thomas Metcalf. Because conjunction is commutative, $P(H\&E) = P(E\&H)$. By the conjunction rule, $P(H\&E) = P(H) \times P(E|H)$, and by that same rule, $P(E\&H) = P(E) \times P(H|E)$. Then we substitute those back in the original equality, yielding $P(H) \times P(E|H) = P(E) \times P(H|E)$, and by a bit of algebra, $P(H|E) = P(E|H) \times P(H) / P(E)$.

[13] In ordinary English, we often use "likelihood" and "probability" interchangeably, but in the Bayesian paradigm, "likelihood" is a technical term, used to mean the probability $P(E|H)$ that the evidence E in question would occur given that hypothesis H (Talbott, 2022, sect. 4.1).

[14] We often find that we initially seem to be missing at least one of these terms, especially the priors. Fortunately, if we have one of the priors, plus two likelihoods, we can usually get the other prior by employing the Law of Total Probability: $P(A) = P(A|B) \times P(B) + P(A|\neg B) \times P(\neg B)$. (Recall that $P(\neg B) = 1 - P(B)$,

so we can get one of those unconditional probabilities if we know the other.) See Weisberg (n.d., pp. 72–73). See n. 15 below for an application.

[15] To calculate probability when we draw from a set without replacement, we use the Hypergeometric Distribution (Weisstein, n.d.). It is possible to run hypergeometric calculations online (Stat Trek, n.d.; Wolfram|Alpha Widgets, 2019). In this example, there is a population of 51, with 3 successes and 48 failures, and a sample size of 10. The probability of getting zero successes in the sample is about 0.511884754 or about 51.2%.

[16] We can use the Hypergeometric Distribution (Weisstein, n.d.) plus some probability calculus (see Introduction to the Probability Calculus by Thomas Metcalf). The Facedown Card has a prior probability of 4/52 of being an ace and of 48/52 of not being an ace. If it is an ace, then there are three successes in the population; if it isn't, then there are four. Given the Hypergeometric Distribution, the probability of drawing zero aces in ten draws if there are four successes in the population is about 0.405. Similarly, if there are three successes, it's about 0.512. Now, given that A and $\neg A$ are mutually exclusive and jointly exhaustive, the probability P(T) then must be equal to P(T&A) + P(T&¬A). The reason is that P(A ∨ ¬A) = 1, so by the rule for conjunctions, P((A ∨ ¬A) & T) = P(T). By elementary logic, ((A ∨ ¬A) & T) is logically equivalent to ((A & T) ∨ (¬A & T)). Then P(T) = P((A & T) ∨ (¬A & T)) (cf. Weisberg, n.d., p. 214). Because the two disjuncts are mutually exclusive, P(T) = P(A&T) + P(¬A & T). In turn, by the rule for conjunctions, P(T&A) + P(T&¬A) = P(T|A) × P(A) + P(T|¬A) × P(¬A) = (4/52) × 0.405 + (48/52) × 0.512 ≈ 41.3%. (See n. 9 above for the derivation of Bayes's Theorem for mutually exclusive, jointly exhaustive hypotheses. See n. 13 above for more explanation. See also Formal Logic: Symbolizing Arguments in Sentential Logic by Thomas Metcalf.)

[17] One prominent example in philosophy is the Fine-Tuning Argument. We have no way to survey the number of worlds in which the universe permits life and count the proportion in which God exists. But we may be able to estimate how confident we should have been in theism before encountering life-permission;

how confident we should be given theism that the universe would permit life; and how confident we should be given atheism that the universe would permit life. Then, using a version of Bayes's Theorem (see n. 9 above), we can estimate the probability given life-permission that God exists, as in P(G|L) = P(L|G)P(G) / [P(L|G)P(G) + P(L|¬G)P(¬G)]. For more on this argument, see <u>The Fine-Tuning Argument for the Existence of God</u> by Thomas Metcalf.

[18] In the sciences, Bayesian reasoning is commonly used to run statistical analyses (Baig, 2020; Kelter, 2020; Tanguy, 2020; cf. Romeijn, 2022). In philosophy, it is commonly used in the philosophy of religion (Draper, 1989; Swinburne, 2004; Collins, 2009) and in decision theory (Jeffrey, 1983; Weisberg, 2022, sect. 6.1). Epistemologists and philosophers of science study Bayesianism and its applications to epistemology and philosophy of science (Joyce, 2005; Talbott, 2022; Weisberg, 2022), including to traditional problems in the philosophy of science (Rinard, 2014).

[19] One dispute between Bayesians is the question of the degree to which we should be subjective or objective Bayesians; see n. 6 above. For examples of puzzles or problems for Bayesianism, including the important Problems of Uncertain Evidence, Old Evidence, and the Priors, see Talbott (2022, sect. 6.2). In turn, the debate over the Principle of Indifference (Keynes, 1921, pp. 52–53) is very relevant to the aforementioned Problem of the Priors (cf. Huemer, 2009, sect. 2; Schupbach, 2022, sect. 2.4; Weisberg, n.d., ch. 18). See Weisberg (n.d., pt. III) for a general account of Bayesianism versus frequentism, and Tanguy (2020) for an argument for the superiority of Bayesianism in hypothesis testing.

References

<u>Baig, Sabeeh A. (2020). Bayesian inference: An introduction to hypothesis testing using Bayes factors.</u> *Nicotine & Tobacco Research, 22*<u>(7), 1244–1246.</u>

Bayes, Mr. and Price, Mr. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London, 53*, 370–418.

Carr, Jennifer Rose. (2013). *Justifying Bayesianism*. [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.

Collins, Robin. (2009). The teleological argument: An exploration of the fine-tuning of the universe. In Craig, W. L. & Moreland, J. P. (eds.), *The Blackwell Companion to Natural Theology* (pp. 202–281). Wiley-Blackwell.

Downey, Allen B. (2012). *Think Bayes: Bayesian Statistics Made Simple*. Green Tea Press.

Draper, Paul. (1989). Pain and pleasure: An evidential problem for theists. *Noûs, 23*(3), 331–350.

Hacking, Ian. (2001). *An Introduction to Probability and Inductive Logic*. Cambridge University Press.

Howson, Colin & Urbach, Peter. (2006). *Scientific Reasoning: The Bayesian Approach* (3rd ed.). Open Court Publishing.

Huber, Franz. (2019). *A Logical Introduction to Probability and Induction*. Oxford University Press.

Huber, Franz. (N.d.). Confirmation and induction. In J. Fieser & B. Dowden (eds.), *The Internet Encyclopedia of Philosophy*.

Huemer, Michael. (2009). Explanationist aid for the theory of inductive logic. *The British Journal for the Philosophy of Science, 60*(2), 345–375.

Huemer, Michael. (2017). There is no pure empirical reasoning. *Philosophy and Phenomenological Research, 95*(3), 592–613.

Jeffrey, Richard C. (1983). *The Logic of Decision* (2nd ed.). University of Chicago Press.

Joyce, James M. (2005). How probabilities reflect evidence. *Philosophical Perspectives, 19*(1), 153–178.

Kelter, Riko. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: a non-technical introduction to Bayesian inference with JASP. *BMC Medical Research Methodology, 20*, Article 142.

Lewis, David K. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability, Volume II* (pp. 263–293). University of California Press.

Rinard, Susanna. (2014). A new Bayesian solution to the Paradox of the Ravens. *Philosophy of Science, 81*(1), 81–100.

Romeijn, Jan-Willem. (2022). Philosophy of statistics. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 ed.).

Steinhart, Eric. (2009). *More Precisely: The Math You Need to Do Philosophy*. Broadview Press.

Schupbach, Jonah N. (2022). *Bayesianism and Scientific Reasoning*. Cambridge University Press.

Stat Trek. (N.d.). Hypergeometric probability calculator. *Stat Trek*.

Swinburne, Richard. (2004). *The Existence of God* (2nd ed.). Oxford University Press.

Talbott, William. (2022). Bayesian epistemology. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 ed.).

Tanguy, Alain. (2020). Hypothesis testing, a simple illustration of the superiority of the Bayesian framework. *Towards Data Science*.

Weisberg, Jonathan. (N.d.). Formal epistemology. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 ed.).

Weisberg, Jonathan. (2011). Varieties of Bayesianism. In D. M. Gabby, S. Hartmann, & J. Woods (eds.), *Handbook of the History of Logic: Volume 10: Inductive Logic* (pp. 477–552). North-Holland.

Weisberg, Jonathan. (2022). Formal epistemology. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 ed.).

Weisstein, Eric W. (N.d.) Hypergeometric distribution. *Wolfram MathWorld*.

Wolfram|Alpha Widgets. (2019). Hypergeometric calculator. *WolframAlpha*.

Vineberg, Susan. (2022). Dutch book arguments. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 ed.).