

**MIT Press • Open Encyclopedia of Cognitive Science**

# **Bayesianism**

**Michael Rescorla<sup>1</sup>**

<sup>1</sup>Department of Philosophy, University of California, Los Angeles

**MIT Press**

**Published on:** Jul 25, 2024

**DOI:** <https://doi.org/10.21428/e2759450.c0cec761>

**License:** [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

*Bayesian decision theory* is a mathematical model of reasoning and decision-making under uncertain conditions. Proponents of Bayesian decision theory are usually called *Bayesians*. Their viewpoint is usually called *Bayesianism*. The Bayesian framework hinges upon two core concepts: *subjective probability* (a numerical measure of the degree to which an agent believes a hypothesis) and *utility* (a numerical measure of how much an agent desires an outcome). Bayesians give rules governing the assignment of probabilities to hypotheses, rules governing how to revise probabilities in light of new evidence, and rules governing how to make decisions in light of probabilities and utilities. Through these rules, Bayesians seek a “mathematization of rationality.” Bayesian decision theory originated as a theory of how rational agents *should* operate, not how they *actually* operate. Recently, cognitive scientists have used it to describe the actual mental activity of humans and some nonhuman animals. *Bayesian cognitive science* offers theories of perception, motor control, navigation, causal reasoning, social cognition, language acquisition, and numerous other psychological domains.

## History

The Bayesian framework rests on insights into probability due to Rev. Thomas Bayes (1763). The framework was first systematically articulated by Pierre-Simon Laplace (1774, 1814/1902), who applied it to astronomy, population statistics, geoscience, jurisprudence, and other domains. The framework assumed its modern form through the work of two thinkers operating independently at around the same time: the philosopher, mathematician, and economist Frank Ramsey (1931) and the statistician Bruno de Finetti (1937/1980).

Among the numerous contributions made by Ramsey and de Finetti, perhaps the most fundamental was their emphasis upon *subjective probability*. Subjective probability measures the degree to which an agent believes a hypothesis. Thus, it is distinct from *objective probability*, which reflects properties of the nonmental world. To illustrate, suppose I flip a coin that is biased to land heads 30% of the time and tails 70% of the time. In a natural sense, the objective chance that the coin will land heads is 30% (or .3 on a scale of 0 to 1). Now suppose that Mary believes the coin is fair rather than biased. Then Mary’s *subjective probability* that the coin will land heads is 50% (or .5 on a scale of 0 to 1). Prior to Ramsey and de Finetti, authors did not distinguish too carefully between subjective and objective probabilities. Ramsey and de Finetti drew the distinction quite sharply, and they insisted that the Bayesian framework should concern itself principally with subjective probabilities.

Building on the work of Ramsey and de Finetti, numerous subsequent thinkers have elaborated and applied Bayesian decision theory. The geophysicist Harold Jeffreys (1961) and the statistician Leonard Savage (1954) developed its conceptual underpinnings. The mathematician A. N. Kolmogorov was not himself a Bayesian, but he supplied the definitive axiomatic treatment of probability (Kolmogorov, 1933/1956). His axiomatization, called *the probability calculus*, provides the canonical mathematical foundation for Bayesian decision theory.

We may distinguish three main ways in which practitioners use Bayesian decision theory:

- *Evaluation*: The goal here is to evaluate how an idealized rational agent should proceed. For example, *game theory* evaluates how agents should engage in strategic interaction ([Harsanyi, 1967](#)), and *formal epistemology* evaluates how agents should fit their opinions to their evidence ([Weisberg, 2009](#)).
- *Guidance*: Here the Bayesian framework serves not to evaluate another (hypothetical) agent but instead to guide *the practitioner*. Two famous examples: Alan Turing and I. J. Good used Bayesian methods during World War II when deciphering the German Enigma code ([Zabell, 2023](#)), and Bayesian analysis located the missing wreckage of Air France 447 after it crashed in the Atlantic Ocean in 2009 ([Stone et al., 2014](#)). On a less dramatic level, scientists routinely use Bayesian inference when comparing rival hypotheses in light of available evidence.
- *Description*: Here one seeks to describe how inference and decision-making *actually* proceed. The system under study may be an agent, a psychological subsystem (such as the perceptual system), or an artifact (such as a computer or robot). The goal is an empirically successful description of how the system makes inferences and decisions.

In practice, the three uses cannot always be disentangled. For example, one reason that the Bayesian framework successfully guides decision-making is that it helps us evaluate how a rational agent would make decisions under similar circumstances. Moreover, a standard justification for describing the human mind through Bayesian models is that humans are at least approximately rational and hence perform fairly well from an evaluative standpoint. Finally, some applications involve a mixture of uses. In machine learning ([Murphy, 2023](#)) and robotics ([Thrun et al., 2005](#)), one often constructs an idealized Bayesian model that evaluates how an agent should behave (e.g., how to navigate through an unfamiliar environment), one then uses the model to guide construction or programming of a physical system, and one then uses the model to describe the physical system.

Beginning in the 1970s, Nobel Prize–winning work by Daniel Kahneman and Amos Tversky cast serious doubt upon whether Bayesian decision theory could yield empirically adequate models of human psychology. Kahneman and Tversky conducted a series of experiments suggesting that humans routinely violate the rules of Bayesian decision theory. A good example is *anchoring bias* ([Tversky & Kahneman, 1974](#)): when people are asked to estimate a quantity (such as the distance between New York City and Boston), their estimates are biased toward a randomly selected number provided to them. Anchoring bias is irrational and violates the rules of Bayesian decision theory. Kahneman and Tversky discovered numerous additional violations. In reaction, many theorists concluded that Bayesian decision theory, while perhaps useful as a model of how one *should* make inferences and decisions, is not useful as a model of how people *do* make inferences and decisions.

The past few decades have witnessed a resurgence of interest among psychologists in descriptive uses of Bayesian decision theory. The resurgence stems from two observations:

- Even if *an agent* violates Bayesian rules, a *psychological subsystem* may conform to the rules. For example, *Bayesian perceptual psychology* holds that the perceptual system estimates environmental conditions through a Bayesian inference based upon proximal sensory stimulations (e.g., retinal stimulations). It is possible that inferences made *by the agent's perceptual system* conform to Bayesian rules even though inferences made *by the agent* violate the rules. Kahneman and Tversky focused on agents rather than psychological subsystems, so their work leaves open the possibility that subsystems are well-described by Bayesian decision theory.
- Even when a system violates Bayesian rules, one can often describe the system as *approximately* satisfying Bayesian rules. For example, [Lieder et al. \(2018\)](#) show that anchoring bias arises naturally from a suitable approximation to idealized Bayesian inference. Thus, a potential role for descriptive Bayesian modeling persists even when agents violate the rules of Bayesian decision theory.

These two observations kindled a proliferation of Bayesian models in cognitive science.

## Core concepts

Subjective probability is commonly notated as  $P(H)$ , where  $H$  is a hypothesis (e.g., the hypothesis that “Seabiscuit will win the race”) and  $P(H)$  is the subjective probability that the agent attaches to  $H$ . Subjective probabilities are measured on a scale from 0 to 1, with 1 being maximal certainty and 0 being utter disbelief. If we are modeling Mary’s subjective probabilities, then the equation

$$P(H) = x$$

means that Mary has subjective probability  $x$  in  $H$ . For example,

$$P(\text{Seabiscuit wins the race}) = \frac{2}{3}$$

means that Mary has subjective probability  $\frac{2}{3}$  that Seabiscuit will win the race. Subjective probabilities are often called *degrees of belief* or *credences*.

Also crucial to the Bayesian framework is the notion of *conditional probability*  $P(H|E)$ : the probability of  $H$  *conditional on*  $E$ . For example, we might consider Mary’s conditional probability

$$P(\text{Seabiscuit wins the race} | \text{Seabiscuit is sick}),$$

i.e. her probability that Seabiscuit will win the race *given that he is sick*. Presumably, this conditional probability is much lower than Mary’s *unconditional* probability that Seabiscuit will win the race. In general,  $P(H|E)$  may differ quite significantly from  $P(H)$ .

## Probabilism

The probability calculus includes three axioms that govern the assignment of probabilities to hypotheses:

- (i) Probabilities fall between 0 and 1. (This axiom sets the scale that we use to measure probabilities.)
- (ii) If a hypothesis is true in every possible outcome, then the hypothesis receives probability 1. For example, if we are investigating whether John has a disease, then the hypothesis that “John either has the disease or does not have the disease” receives probability 1 because it is true in every possible outcome.
- (iii) If  $A$  and  $B$  are mutually exclusive hypotheses, then

$$P(A \text{ or } B) = P(A) + P(B).$$

For example,

$$\begin{aligned} &P(\textit{Seabiscuit wins the race or War Admiral wins the race}) \\ &= P(\textit{Seabiscuit wins the race}) + P(\textit{War Admiral wins the race}), \end{aligned}$$

because Seabiscuit and War Admiral cannot both win the race. This axiom is called *additivity*.

Axioms (i)–(iii) can be formulated in more rigorous terms, but the present formulations suffice for our purposes.

Using the axioms, one can prove theorems about probability. For example, let  $B$  and  $C$  be mutually exclusive and jointly exhaustive hypotheses: either  $B$  is true or  $C$  is true, and they are not both true. The *law of total probability* holds that

$$P(A) = P(A|B)P(B) + P(A|C)P(C).$$

In words:  $A$ 's probability can be found by adding together its probability *conditional on*  $B$  times the probability of  $B$  and its probability *conditional on*  $C$  times the probability of  $C$ .

A core Bayesian thesis is that credences (i.e., subjective probabilities) should conform to the probability calculus axioms. The thesis is sometimes called *probabilism*. To endorse probabilism is not to claim that credences always *do* satisfy the probability calculus axioms. It is just to claim that rationality demands conformity. For example, an agent's credences might violate the additivity axiom (iii), and in that case probabilism castigates the agent as irrational.

One classic argument for probabilism hinges upon *gambling*. The argument holds that, when an agent's credences violate the axioms, a cunning opponent can induce the agent to accept a series of bets that inflict a *sure loss* ([de Finetti, 1937/1980](#); [Ramsey, 1931](#)). The agent regards the bets as fair yet loses money in every

possible outcome. Vulnerability to a sure loss is taken as indicative of irrationality: if you regard a series of bets as fair even though those bets are guaranteed to lose you money, then something must have gone awry in your reasoning. Although this argument is still popular, it has been attacked from several angles ([Hájek, 2009](#)). In particular, some theorists question whether vulnerability to a sure loss is necessarily indicative of irrationality.

## Updating subjective probability

Suppose I begin with subjective probability  $P(H)$  and then receive evidence  $E$ . What new probability should I assign to  $H$ ? Bayesians hold that I should adopt the conditional probability  $P(H|E)$  as my new *unconditional* probability in  $H$ .  $P(H|E)$  is the probability of  $H$  conditional on  $E$ , and I have learned that  $E$  is true, so I should replace  $P(H)$  with  $P(H|E)$  as my new credence in  $H$ . For example, if I learn that Seabiscuit is sick, then I should adopt my former *conditional* probability

$$P(\text{Seabiscuit wins the race} | \text{Seabiscuit is sick})$$

as my new *unconditional* probability that Seabiscuit will win the race. When a Bayesian agent updates credences in this way, the agent is said to *conditionalize* on  $E$ .

Why should agents conditionalize? One popular argument, echoing the classic argument for probabilism, is that any update rule other than conditionalization leaves the agent vulnerable to a sure loss ([Lewis, 1999](#); [Rescorla, 2018](#); [Skyrms, 1987](#)). This argument, like the parallel argument for probabilism, comes under attack from various directions ([Hájek, 2009](#)).

$P(H)$  is called the *prior probability*.  $P(H|E)$  is called the *posterior probability*. The posterior can be computed using *Bayes's theorem*:

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}.$$

$P(E|H)$ , called the *prior likelihood*, is the conditional probability of  $E$  given  $H$ .  $P(E|H)$  is higher when the hypothesis  $H$  renders the evidence  $E$  more likely. The denominator  $P(E)$  is lower when the evidence is relatively surprising or unexpected. Thus, Bayes's theorem says that  $H$  receives higher posterior probability to the extent that it has higher prior probability, to the extent that it renders evidence  $E$  more likely, and to the extent that evidence  $E$  is relatively surprising. Intuitively, evidence raises the probability of a hypothesis to the extent that the evidence is unexpected but *would be expected assuming the hypothesis*.

As a standard illustration of Bayes's theorem, suppose John wonders whether he has some disease that afflicts 2.5% of the general population. He takes a blood test that has a false positive rate of 5% and a false negative rate of 5%. Let *Disease* be the statement that John has the disease, *No Disease* the statement that he does not

have the disease, and *Positive* the statement that the test result is positive. Assume that John pegs his prior  $P(Disease)$  to the base rate of the disease in the general population, setting

$$P(Disease) = .025.$$

Assume that he pegs his prior likelihoods  $P(Positive|Disease)$  and  $P(Positive|No Disease)$  to the rates of false positives and negatives, setting

$$P(Positive|Disease) = .95$$

and

$$P(Positive|No Disease) = .05.$$

Suppose John receives a positive result. What should he adopt as his new probability that he has the disease? By Bayes's theorem,

$$P(Disease|Positive) = \frac{P(Disease)P(Positive|Disease)}{P(Positive)}.$$

By the law of total probability,

$$P(Positive) = P(Positive|Disease)P(Disease) + P(Positive|No Disease)P(No Disease) = .95$$

$P(Positive)$  is low because John probably does not have the disease (few people do) and so will probably not test positive. Plugging these numbers into Bayes's theorem:

$$P(Disease|Positive) = \frac{.025 \times .95}{.0725} \approx .3276.$$

The base rate of the disease in the general population is so low that, even though John received a positive test and even though the test is fairly accurate, he still probably does not have the disease. If John were to neglect the base rate and conclude from his positive test result that he probably has the disease, then he would commit *the base rate fallacy*.

This example illustrates how Bayes's theorem helps agents integrate new evidence with prior information: John combines his new evidence (the positive test result) with his prior information (the base rate of the disease) to update his probability that he has the disease. In more complicated cases, many additional pieces of evidence may be available. For example, John may display symptoms of the disease, or he may learn that he has a gene that correlates with the disease, and so on. Bayes's theorem provides a general basis for statistical inference sparked by incoming evidence.

## Expected utility

If  $C$  is a possible outcome, then  $U(C)$  is the *utility* that an agent assigns to  $C$ .  $U(C)$  is greater than  $U(D)$  when the agent regards  $C$  as more desirable than  $D$ . Roughly speaking, utility measures how strongly the agent wishes an outcome to occur. Utility does not necessarily measure how *happy* an outcome will make the agent. I might regard  $C$  as more desirable than  $D$  even though  $C$  will make me much less happy. For example, I may regard  $C$  as better than  $D$  from the perspective of what is fair, or honorable, and so on, even though  $C$  will leave me much less happy than  $D$ .

When making decisions, it is usually uncertain which outcome will result from an action. For example, if I decide to go on a picnic, then it is uncertain whether I will get wet because it is uncertain whether it will rain. According to Bayesians, the key notion for making decisions is *expected utility*, the utility I expect from performing an action. Take the simplest case, where there are two possible outcomes  $C$  and  $D$  from performing action  $A$ . The expected utility of  $A$ —notated  $EU(A)$ —may be defined as

$$EU(A) = P(C|A)U(C) + P(D|A)U(D).$$

$EU(A)$  is a weighted average of the possible utilities that might result from action  $A$ , where weights are given by probabilities conditional on my performing  $A$ . For example, let  $A$  be the action of going on a picnic,  $C$  the outcome that it rains while I am on the picnic, and  $D$  the outcome that it does not rain while I am on the picnic. Assuming I would prefer not to get wet,  $EU(A)$  will be relatively high when I have a low subjective probability of rain, relatively low when I have a high subjective probability of rain. The less likely it is to rain, the higher my expected utility of picnicking. This treatment extends to cases where my action has more than two possible outcomes.

There are alternative possible definitions of expected utility ([Briggs, 2023](#)). The alternative definitions share the same core intuitive idea: expected utility is a weighted average of the utilities that might result from action  $A$ , where each weight measures the likelihood of an outcome given that the agent has performed  $A$ .

A key tenet of Bayesian decision theory is that rational agents should maximize expected utility. Suppose I am choosing among actions  $A_1, A_2, \dots, A_n$ . According to Bayesians, I should compare the expected utilities  $EU(A_1), EU(A_2), \dots, EU(A_n)$  and pick whichever action has the highest expected utility. I thereby pick whichever action I expect will lead to the highest possible utility under my current circumstances. Expected utility maximization is a linchpin of economics and many other applications of the Bayesian framework.

## Bayesian modeling of the mind

Bayesian cognitive science constructs descriptive models of mental activity [see [Bayesian Models of Cognition](#)] ([Chater & Oaksford, 2008](#); [Griffiths et al., 2008](#); [Ma et al., 2023](#)).

We posit that a psychological system conforms (at least approximately) to the rules of Bayesian decision theory. We then try to isolate the prior probability  $P(H)$  and the prior likelihood  $P(E|H)$  employed by the system so that we can describe the system as computing the posterior  $P(H|E)$  or an approximation thereof. Typically, these Bayesian computations are inaccessible to conscious introspection. For example, Bayesian perceptual psychology postulates that the perceptual system executes a Bayesian inference based upon proximal sensory stimulations. *The perceiver* does not execute the inference and is not aware of it. No matter how hard you try, you cannot introspect the Bayesian inferences executed by your perceptual system.

$H$  and  $E$  vary with the psychological task being modeled. If we are modeling perception ([Knill & Richards, 1996](#); [Rescorla, 2015](#)), then  $H$  concerns observable properties of perceived objects (e.g., shape, size, color, location, etc.), and  $E$  concerns proximal sensory stimulations (e.g., retinal stimulations) through which the perceptual system estimates observable properties. If we are modeling navigation ([Chen et al., 2017](#); [Lakshminarasimhan et al., 2018](#)), then  $H$  concerns self-location and the locations of salient landmarks, and  $E$  concerns navigational cues such as optic flow, vestibular information, or egocentric distances to landmarks. If we are modeling language acquisition ([Pearl, 2021, 2023](#)), then  $H$  concerns possible grammars for the language, and  $E$  concerns possible utterances encountered by the language learner. If we are modeling how the agent interprets a second agent's behavior ([Baker et al., 2005](#)), then  $H$  concerns the second agent's goals (e.g., "Drew wants to turn on the light"), and  $E$  concerns the second agent's observed behavior (e.g., "Drew is walking towards the light switch").

$P(H)$  encodes prior information about the world. In Bayesian perceptual psychology, for example, it is common to posit prior probabilities that favor slow speeds, regular surface textures, or an overhead lighting direction.  $P(E|H)$  encodes prior information about the interface between mind and world. In Bayesian perceptual psychology, it typically reflects physical laws governing the interaction between the perceiver and the environment (e.g., laws of optics govern how light travels from the perceived object to the retina). Assignment of unconditional and conditional probabilities is not the only way to encode prior information. In some cases, the psychological system may encode prior information by delimiting the hypothesis space (i.e., assigning probabilities only to certain hypotheses and eliminating other hypotheses from consideration altogether). For instance, Bayesian modeling of language acquisition typically assumes a restricted hypothesis space that recognizes only certain grammars as possible options.

Once we have articulated a Bayesian model of a psychological task, we can ask how well the model fits actual performance. How well does a Bayesian model of speed perception fit human speed perception? How well does a Bayesian model of language acquisition fit human language acquisition? Such questions can be investigated through normal scientific procedures, by comparing the model's predictions with experimental data. Empirical tests in this vein suggest that, in many cases, Bayesian models fit actual performance quite well.

## Questions, controversies, and new developments

The Bayesian framework is popular but controversial. Critics raise many doubts about the framework, such as whether expected utility maximization is an adequate model of rational decision-making (e.g., [Allais, 1953](#); [Buchak, 2013](#); [Ellsberg, 1961](#)), whether agents should update their credences through conditionalization (e.g., [Carr, 2015](#); [Levi, 1980](#); [Williamson, 2002](#)), and whether agents have precise credences in the first place (e.g., [Kyburg, 1987](#); [Levi, 1980](#); [Shafer, 1976](#); [Walley, 1991](#)). Among Bayesians, internal disputes arise over how to define expected utility ([Briggs, 2023](#)), how to justify probabilism and conditionalization (e.g., [Hájek, 2008](#); [Joyce, 1998](#)), whether to revise Kolmogorov's specific formulation of the probability calculus axioms (e.g., [Howson, 2014](#); [Lewis, 1980](#); [Liu, 2020](#); [Lyon, 2016](#); [Skyrms, 1980](#)), and other matters. We will confine attention to a few notable issues.

### *Bayesian statistics and classical statistics*

Statisticians usually contrast Bayesianism with an alternative framework called *classical statistics* ([Fisher, 1956](#); [Neyman & Pearson, 1967](#)). Whereas Bayesian statistics employs a subjectivist conception of probability, classical statistics employs a *frequentist* conception. Frequency is a type of objective probability. On the frequentist conception, the probability of *A* is (roughly speaking) the long-run frequency with which *A* would occur in a repeated experiment. For example, the probability of a coin landing heads is the long-run frequency with which the coin would land heads if flipped repeatedly.

Bayesian statistics and classical statistics enshrine fundamentally different perspectives on statistical inference. Classical statisticians allege that the Bayesian framework is unscientific due to its reliance on subjective probability. Bayesians respond that statistical inference inevitably begins from a subjective starting point, just as logical inference begins with premises accepted by the reasoner. Bayesian statisticians make all subjective elements explicit through the priors. Classical statisticians leave subjective elements implicit in their inferential procedures. For example, classical statisticians say that a result is *statistically significant* when it has  $p < .05$ , yet this threshold looks subjective and arbitrary (why .05 rather than .049 or .048 or .005?). According to Bayesians, it is better to make subjective elements as explicit as possible.

A common anti-frequentist argument is that there are situations where we want to assign probabilities even though frequencies look inapplicable. To illustrate, suppose I am betting on the outcome of a future presidential election. Then I need to assess how probable it is that each candidate will win the election. Frequencies are irrelevant: there is no meaningful prospect of repeating the election many times and seeing who wins each time. The operative conception of probability is subjective rather than frequentist: my betting choices reflect how probable *I* deem it that each candidate will win. I also need rules for updating my probabilities in light of new evidence, such as new polling data or a major scandal. These considerations push us toward the Bayesian framework for at least some probabilistic applications.

The dispute between Bayesian and classical statisticians has raged for almost a century. It still finds vocal advocates on either side and shows no sign of abating ([Sprenger, 2016](#)).

### ***The problem of the priors***

Bayesian inference presupposes a prior probability  $P(H)$  and a prior likelihood  $P(E|H)$ . Which priors should one use? A long tradition seeks to answer this question by articulating rational constraints upon priors. Most famously, the *Principle of Indifference* holds that, absent further information, each alternative hypothesis should receive equal probability ([Carnap, 1950](#); [Jaynes, 1973](#); [Laplace, 1814/1902](#); [Williamson, 2007](#)). Another popular strategy is to constrain priors by citing frequencies or objective chances ([Lewis, 1980](#)). These suggestions have received extensive discussion, some supportive and some critical.

The problem of the priors assumes a different form in the context of Bayesian cognitive science. The question here is how the mind comes to have the priors that it actually has. If we postulate that perceptual priors favor slower speeds, then we must ask how the perceptual system came to favor slower speeds. If we posit that Bayesian language acquisition eliminates certain grammars from consideration, then we naturally wonder how those grammars were eliminated. How do nature and nurture interact to shape the hypothesis space and the priors? Since perceptual priors change rapidly in response to changing environmental conditions ([Adams et al., 2004](#); [Petzschner et al., 2015](#)), it is clear that the agent's past experience sometimes plays an important role. In other cases, such as language acquisition, priors may be due entirely or almost entirely to genetic heritage.

### ***Opposition to Bayesian cognitive science***

Critics raise a number of challenges to Bayesian cognitive science. Some critics maintain that, when a Bayesian model fails to specify how the priors arise, it is fatally incomplete ([Orlandi, 2014](#)). Other critics worry that the Bayesian research program is vacuous or unfalsifiable ([Bowers & Davis, 2012](#)). They complain that, when a Bayesian model does not fit experimental data, one can simply alter the priors to achieve a better fit. They conclude that the Bayesian program does not have the testable content one expects from a genuine scientific hypothesis. Another criticism targets *mechanisms* ([Jones & Love, 2011](#)). When we model a psychological system in Bayesian terms, we do not describe underlying algorithms or neural processes through which the system converts priors into the posterior. So Bayesian models have a relatively nonmechanistic character when compared with many models found in psychology and neuroscience. Critics charge that this nonmechanistic character renders Bayesian models unexplanatory. For responses to these and other criticisms of Bayesian cognitive science, see ([Griffiths et al., 2012](#); [Rescorla, 2020](#)).

### ***Neural implementation***

Although Bayesian models of the mind do not specify underlying neural implementation mechanisms, proponents agree that we should try to illuminate those mechanisms. How are the priors and the posterior physically realized in the brain? Through what neural processes does the brain convert priors into the

posterior? These questions are under active investigation within computational neuroscience ([Fiser et al., 2010](#); [Pouget et al., 2013](#); [Rescorla, 2024](#)). Researchers have offered detailed theories of how the brain might implement (or approximately implement) Bayesian inference. Some neural implementation theories fit fairly well with neurophysiological data, though research on this topic is still at a relatively early stage.

Current neural implementation theories agree upon a crucial point: Subjective probabilities need not be explicitly enumerated in the brain. For example, if a psychological system assigns probability .6 to a hypothesis, the number .6 need not be explicitly notated or represented. Indeed, explicit enumeration is usually not even possible: There are usually infinitely many distinct hypotheses that receive probability assignments, and it is not possible to enumerate infinitely many distinct probability assignments. Current neural implementation theories instead postulate that neural states *implicitly encode* probabilities. The question then becomes how implicit encoding works and how neural processes exploit it to implement (or approximately implement) Bayesian inference.

## Broader connections

Bayesian decision theory stands at the nexus of science, philosophy, mathematics, and praxis. The Bayesian framework engages a host of foundational questions regarding rationality, evidence, inference, and decision-making ([Earman, 1992](#); [Weisberg, 2009](#)). In many instances, these questions can be illuminated or at least made more precise through rigorous Bayesian analysis. The framework also bears upon any project that requires formal modeling of inference and decision-making, so it has countless applications stretching across a vast range of undertakings, including but not limited to finance, medicine, government, and industry.

When used to construct descriptive psychological models, the framework connects with diverse topics of interest to contemporary cognitive science. Bayesian psychological modeling is a species of the (controversial) *rational analysis* research program, which seeks to convert evaluative models of cognition into empirically adequate descriptions ([Anderson, 1990](#)). Bayesian modeling also raises questions about how a mind with limited computational resources at its disposal can best use those resources so as to approximate the rational ideal ([Icard, 2018](#); [Lieder & Griffiths, 2019](#)). On a different note, the problem of the priors intersects with longstanding debates concerning how much human cognition is innate ([Scholl, 2005](#)), such as debates regarding the innateness of linguistic knowledge ([Pearl, 2021](#)) and the innateness of concepts ([Piantadosi et al., 2012](#)) [see [Language Acquisition](#)]. Finally, Bayesian modeling presupposes an ability to represent the world. For example, an agent can have subjective probability .6 that Seabiscuit will win the race only if the agent can represent the possibility that Seabiscuit wins the race. Thus, Bayesian modeling harmonizes with widespread cognitive science appeals to mental representation ([Fodor, 1975](#); [Rescorla, in press](#)). As the Bayesian program develops, it may shed further light upon these and other aspects of the mind.

## Further reading

- McGrayne, S. (2011). *The theory that would not die*. Yale University Press.

- Ma, W. J., Kording, K., & Goldreich, D. (2023). *Bayesian models of perception and action: An introduction*. MIT Press.
- Rescorla, M. (in press). *Bayesian models of the mind*. Cambridge University Press.
- Weisberg, J. (2009). Varieties of Bayesianism. In D. M. Gabbay, S. Hartman, & J. Woods (Eds.), *Handbook of the history of logic: Inductive logic* (Vol. 10, pp. 477–551). Elsevier.

## References

- Adams, W., Graf, E., & Ernst, M. (2004). Experience can change the ‘light-from-above’ prior. *Nature Neuroscience*, 7, 1057–1058. <https://doi.org/10.1038/nn1312>
- Allais, M. (1953). Le comportement de l’homme rationnel devant le risque: Critique des postulats et axiomes de l’école américaine. *Econometrica*, 21(4), 503–546. <https://doi.org/10.2307/1907921>
- Anderson, J. (1990). *The adaptive character of thought*. Lawrence Erlbaum.
- Baker, C., Tenenbaum, J., & Saxe, R. (2005). Bayesian models of human action understanding. In Y. Weiss, B. Schölkopf & J. Platt (Eds.), *Advances in Neural Information Processing Systems 18 (NIPS 2005)* (pp. 99–106). MIT Press.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 470–418. <https://doi.org/10.1098/rstl.1763.0053>
- Bowers, J., & Davis, C. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3), 389–414. <https://doi.org/10.1037/a0026450>
- Briggs, R. (2023). Normative theories of rational choice: Expected utility. In E. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2023 Edition). <https://plato.stanford.edu/archives/win2023/entries/rationality-normative-utility/>
- Buchak, L. (2013). *Risk and rationality*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199672165.001.0001>

- Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press.

↑

- Carr, J. (2015). Don't stop believing. *Canadian Journal of Philosophy*, 45(5-6), 744–766.  
<https://doi.org/10.1080/00455091.2015.1123454>

↑

- Chater, N., & Oaksford, M. (Eds.). (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199216093.001.0001>

↑

- Chen, X., McNamara, T., Kelly, J., & Wolbers, T. (2017). Cue combination in human spatial navigation. *Cognitive Psychology*, 95, 105–144. <https://doi.org/10.1016/j.cogpsych.2017.04.003>

↑

- de Finetti, B. (1980). Foresight: Its logical laws, its subjective sources. In H. Kyburg, Jr. & H. Smokler (Eds.), *Studies in subjective probability* (2nd ed., pp. 93–158). Robert E. Krieger. (Original work published 1937)

↑

- Earman, J. (1992). *Bayes or bust?* MIT Press.

↑

- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, 75(4), 643–669. <https://doi.org/10.2307/1884324>

↑

- Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences*, 14(3), 119–130.  
<https://doi.org/10.1016/j.tics.2010.01.003>

↑

- Fisher, R. A. (1956). *Statistical methods and scientific inference*. Hafner.

↑

- Fodor, J. (1975). *The language of thought*. Thomas Y. Crowell.

↑

- Griffiths, T., Chater, N., Norris, D., & Pouget, A. (2012). How Bayesians got their beliefs (and what those beliefs really are): Comment on Bowers and Davis (2012). *Psychological Bulletin*, 138(3), 415–422.  
<https://doi.org/10.1037/a0026884>

[↵](#)

- Griffiths, T., Kemp, C., & Tenenbaum, J. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59–100). Cambridge University Press.

[↵](#)

- Hájek, A. (2008). Arguments for—or against—probabilism? *The British Journal for the Philosophy of Science*, 59(4), 793–819. <https://doi.org/10.1093/bjps/axn045>

[↵](#)

- Hájek, A. (2009). Dutch book arguments. In P. Anand, P. Pattanaik, & C. Puppe (Eds.), *The handbook of rationality and social choice* (pp. 173–195). Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780199290420.003.0008>

[↵](#)

- Harsanyi, J. (1967). Games with incomplete information played by “Bayesian” players, I-III. Part I: The basic model. *Management Science*, 14(3), 159–182. <https://doi.org/10.1287/mnsc.14.3.159>

[↵](#)

- Howson, C. (2014). Finite additivity, another lottery paradox, and conditionalization. *Synthese*, 191, 989–1012. <https://doi.org/10.1007/s11229-013-0303-3>

[↵](#)

- Icard, T. (2018). Bayes, bounds, and rational analysis. *Philosophy of Science*, 85(1), 79–101.  
<https://doi.org/10.1086/694837>

[↵](#)

- Jaynes, E. T. (1973). The well-posed problem. *Foundations of Physics*, 3, 477–493.  
<https://doi.org/10.1007/BF00709116>

[↵](#)

- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

[↵](#)

- Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(4), 169–188. <https://doi.org/10.1017/S0140525X10003134>

[↵](#)

- Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4), 575–603.  
<https://doi.org/10.1086/392661>

↑

- Knill, D., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511984037>

↑

- Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (2nd English ed., N. Morrison, Ed. & Trans.) Chelsea Publishing Company. (Original work published 1933)

↑

- Kyburg, H. (1987). Bayesian and non-Bayesian evidence and updating. *Artificial Intelligence*, 31(3), 271–293. [https://doi.org/10.1016/0004-3702\(87\)90068-3](https://doi.org/10.1016/0004-3702(87)90068-3)

↑

- Lakshminarasimhan, K. J., Petsalis, M., Park, H., DeAngelis, G. C., Pitkow, X., & Angelaki, D. E. (2018). A dynamic Bayesian observer model reveals origins of bias in visual path integration. *Neuron*, 99(1), 194–206.e5. <https://doi.org/10.1016/j.neuron.2018.05.040>

↑

- Laplace, P. S. (1774). Memoire sur les suites recurro-recurrentes et sur leurs usages dans la theorie des hasards. *Memoires de Mathematique et de Physique Presentes á l'Academie Royale des Sciences, par divers Savans, and Luts dans ses Assemblies*, 6, 353–371.

↑

- Laplace, P.-S. (1902). *A philosophical essay on probabilities* (F. W. Truscott & F. L. Emory, Trans.). Wiley. (Original work published 1814)

↑

- Levi, I. (1980). *The enterprise of knowledge: An essay on knowledge, credal probability, and chance*. MIT Press.

↑

- Lewis, D. (1980). A subjectivist's guide to objective chance. In R. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 2, pp. 263–293). University of California Press. <https://doi.org/10.1525/9780520318328-009>

↑

- Lewis, D. (1980). A subjectivist's guide to objective chance. In R. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 2). University of California Press. <https://doi.org/10.1525/9780520318328-009>

↑

-

[↵](#)

- Lieder, F., & Griffiths T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43(1), Article e1. <https://doi.org/10.1017/S0140525X1900061X>

[↵](#)

- Lieder, F., Griffiths, T. L., Huys, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, 25, 322–349. <https://doi.org/10.3758/s13423-017-1286-8>

[↵](#)

- Liu, Y. (2020). Countable additivity, idealization, and conceptual realism. *Economics and Philosophy*, 36(1), 127–147.

[↵](#)

- Lyon, A. (2016). Kolmogorov’s axiomatization and its discontents. In A. Hájek & C. Hitchcock (Eds.), *The Oxford handbook of the philosophy of probability* (pp. 155–166). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199607617.013.9>

[↵](#)

- Ma, W. J., Kording, K., & Goldreich, D. (2023). *Bayesian models of perception and action: An introduction*. MIT Press.

[↵](#)

- Murphy, K. P. (2023). *Probabilistic machine learning: Advanced Topics*. MIT Press.

[↵](#)

- Neyman, J., & Pearson, E. S. (1967). *Joint statistical papers*. Cambridge University Press. <https://doi.org/10.1525/9780520339897>

[↵](#)

- Orlandi, N. (2014). *The innocent eye: Why vision is not cognitive process*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199375035.001.0001>

[↵](#)

- Pearl, L. S. (2021). How statistical learning can play well with universal grammar. In N. Allott, T. Lohndal, & G. Rey (Eds.), *A companion to Chomsky* (pp. 267–286). Wiley. <https://doi.org/10.1002/9781119598732.ch17>

[↵](#)

-

↑

- Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, 19(5), 285–293. <https://doi.org/10.1016/j.tics.2015.03.002>

↑

- Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, 123(2), 199–217. <https://doi.org/10.1016/j.cognition.2011.11.005>

↑

- Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience*, 16, 1170–1178. <https://doi.org/10.1038/nn.3495>

↑

- Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–199). Routledge and Kegan Paul.

↑

- Rescorla, M. (2015). Bayesian perceptual psychology. In M. Matthen (Ed.), *The Oxford handbook of the philosophy of perception* (pp. 694–716). Oxford University Press.

↑

- Rescorla, M. (2018). A Dutch book theorem and converse dutch book theorem for Kolmogorov conditionalization. *The Review of Symbolic Logic*, 11(4), 705–735. <https://doi.org/10.1017/S1755020317000296>

↑

- Rescorla, M. (2020). A realist perspective on Bayesian cognitive science. In A. Nes & T. Chan (Eds.), *Inference and consciousness* (pp. 40–73). Routledge.

↑

- Rescorla, M. (2024). Neural implementation of (approximate) Bayesian inference. In T. Cheng, R. Sato, & J. Hohwy (Eds.), *Expected experiences: The predictive mind in an uncertain world* (pp. 197–239). Routledge.

↑

- Rescorla, M. (in press). *Bayesian models of the mind*. Cambridge University Press.

↑

- Savage, L. (1954). *The foundations of statistics*. Wiley.

↑

- Scholl, B. (2005). Innateness and (Bayesian) visual perception: Reconciling nativism and development. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 34–52). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195179675.003.0003>

↵

- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press. <https://doi.org/10.1515/9780691214696>

↵

- Skyrms, B. (1980). *Causal necessity: A pragmatic investigation of the necessity of laws*. Yale University Press.

↵

- Skyrms, B. (1987). Dynamic coherence and probability kinematics. *Philosophy of Science*, 54(1), 1–20. <https://doi.org/10.1086/289350>

↵

- Sprenger, J. (2016). Bayesianism vs. frequentism in statistical inference. In A. Hájek & C. Hitchcock (Eds.), *The Oxford handbook of the philosophy of probability* (pp. 382–405). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199607617.013.23>

↵

- Stone, L., Keller, C., Kratzke, T., & Strumpfer, J. (2014). Search for the wreckage of Air France Flight AF 447. *Statistical Science*, 29(1), 69–80. <https://doi.org/10.1214/13-STS420>

↵

- Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic robotics*. MIT Press.

↵

- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124–1131.

↵

- Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman and Hall.

↵

- Weisberg, J. (2009). Varieties of Bayesianism. In D. M. Gabbay, S. Hartman, & J. Woods (Eds.), *Handbook of the history of logic: Inductive logic* (Vol. 10, pp. 477–551). Elsevier.

↵

-

↩

- Williamson, T. (2002). *Knowledge and its limits*. Oxford University Press.  
<https://doi.org/10.1093/019925656X.001.0001>

↩

- Zabell, S. (2023). The secret life of I. J. Good. *Statistical Science*, 38(2), 285–302.  
<https://doi.org/10.1214/22-STS870>

↩