MIT Press • Open Encyclopedia of Cognitive Science

Bayesian Models of Cognition

Thomas L. Griffiths[1]

[1]Princeton University

MIT Press

Published on: Jul 24, 2024

DOI: https://doi.org/10.21428/e2759450.7b420317

Bayesian models of cognition explain aspects of human behavior as a result of rational probabilistic inference. In particular, these models make use of Bayes' rule, which indicates how rational agents should update their beliefs about hypotheses in light of data. Bayes' rule provides an optimal solution to inductive problems for which the observed data are insufficient to distinguish between hypotheses. Since many of the things that human minds need to do involve inductive problems——from identifying the structure of the world based on limited sensory data to inferring what other people think based on their behavior——Bayesian models have broad applicability within cognitive science. Being able to identify what a rational agent would do in these situations provides a way to explain why people might act similarly, and is a tool for exploring the implicit assumptions underlying human behavior. In particular, Bayesian models make it easy to explore the inductive biases that inform people's inferences, being those factors other than the data that guide people in selecting one hypothesis over another.

History

Bayesian models of cognition are based on a subjective interpretation of probability originally introduced in the 18th century by Thomas Bayes and Pierre-Simon Laplace [see Bayesianism]. In this approach, probabilities are taken to indicate the degree of belief that an agent assigns to an event. Probability theory then specifies how these degrees of belief should behave. Bayes' rule is just a basic result in probability theory, but the subjective interpretation of probability turns it into a powerful tool for understanding how rational agents should update their degrees of belief based on evidence.

Bayesian models of human cognition were first proposed in the 1950s as mathematical modeling began to come into contact with behavioral experiments during the cognitive revolution. In particular, subjective probability and Bayes' rule were evaluated as an account of human cognition when ideas from statistical decision theory were first used to study human decision-making (see Edwards, 1961). Early research comparing human belief updating to the prescriptions of probability theory concluded that people acted in a way that was consistent with Bayesian inference but were "conservative," updating their beliefs more slowly than indicated by Bayes' rule (Phillips & Edwards, 1966; Peterson & Beach, 1967).

The Bayesian approach ran into a significant challenge in the work of Kahneman and Tversky (1972), who showed that people deviated systematically from probability theory when assessing the probability of events. In particular, people could be induced to focus on the evidence provided to them rather than combining this evidence with base rates in the way that Bayes' rule stipulates. Kahneman and Tversky (1972) concluded that "For anyone who would wish to view man as a reasonable intuitive statistician, such results are discouraging." (p. 445).

The discouragement lasted until the late 1980s, when Roger Shepard developed a different way to use probability theory to model human cognition. Rather than focusing on probability judgment, Shepard was

interested in explaining how people form generalizations. He realized that this problem could be cast as one of probabilistic inference—assessing the probability that a property of one object is shared by another—and used Bayes' rule to derive the optimal solution to this problem (Shepard, 1987). In doing so, he was seeking to drive a universal law—a principle that should hold for any intelligent organism anywhere in the universe. By indicating how a rational agent should solve this cognitive problem, Bayes' rule provided an explanation for the patterns seen in the generalization behavior of humans and other animals.

John Anderson took inspiration from Shepard's approach of using probability theory to study a wider range of cognitive processes and developed it into the framework of rational analysis (Anderson, 1990). In this framework, aspects of human cognition are explained by identifying the problem that the human mind has to solve and then deriving the optimal solution to that problem. If the optimal solution is consistent with human behavior, it provides a way to understand why people behave in the way that they do. For problems that involve inductive inference, in which the observed data do not provide enough information to be certain about the process that produced it, Bayes' rule provides an optimal solution. As a consequence, Anderson (1990) developed Bayesian models of categorization, memory, causal learning, and problem solving. This in turn inspired Oaksford and Chater (1994) to show that a classic fallacy of logical reasoning could be explained in terms of Bayesian updating and optimal information seeking, demonstrating that evidence for human irrationality could be reinterpreted using this approach.

The 21st century has seen the broad application of Bayesian models to different aspects of human cognition, spurred by the development of new methods for statistical inference and probabilistic reasoning in statistics and machine learning. These aspects of human cognition include concept learning (Tenenbaum, 2000), word learning (Xu & Tenenbaum, 2007), causal learning (Griffiths & Tenenbaum, 2005; Lu et al., 2008), predicting the future (Griffiths & Tenenbaum, 2006), motor control (Körding & Wolpert, 2006), theory of mind (Baker et al., 2009; Jara-Ettinger et al., 2016), categorization (Goodman et al., 2008; Sanborn et al., 2010), speech perception (Feldman et al., 2009), language learning (Perfors et al., 2011), word recognition (Norris, 2006), and intuitive physics (Sanborn et al., 2013; Battaglia et al., 2013).

Core concepts

The heart of Bayesian models of cognition is Bayes' rule, which makes it possible to learn structured representations, explore human inductive biases, and perform probabilistic inference using rich generative models.

Bayes' rule

Bayes' rule indicates how agents should update their degrees of belief in hypotheses given observed data. Assume that a hypothesis h from a set of hypotheses H is under consideration. The degree of belief assigned to the hypothesis before observing any data is $P(h)$, known as the prior probability. After observing data d, the degree of belief assigned to the hypothesis $P(h|d)$ is called the posterior probability (the | symbol should be

read as "given," so this is the probability of *h* given, or taking into account, the information contained in *d*). Bayes' rule applies the definition of the conditional probability from probability theory to give

$$P(d) = \frac{P(h)P(h)}{\sum_{h' \in H} P(h')\,P(h')}$$

where $P(d|h)$ is the probability of observing *d* if *h* were true, known as the likelihood. The sum in the denominator simply adds up the same quantity (the product of the prior probability and the likelihood) over all of the hypothesis in *H*, making sure that the posterior probability $P(h|d)$ sums to 1 over all hypotheses. The numerator is thus the key to Bayes' rule, indicating that how much we believe in a hypothesis after seeing data should reflect the product of the prior probability of that hypothesis and the probability of the data if that hypothesis were true.

Intuitively, Bayes' rule says that our beliefs about hypotheses should be a function of two factors: how plausible those hypotheses are (as reflected in the prior probability) and how well they fit the observed data (as reflected in the likelihood). These two factors contribute equally, and do so multiplicatively—if either one of them is very small, the other has to be very large to compensate. As a simple example, imagine looking out the window during the summer and seeing gray clouds (the data *d*). You might consider three hypotheses: that the day will be sunny, that it will rain, and that there is a nearby forest fire. Sunny days might be more frequent than rainy days, which are more frequent than days where there are forest fires, so the prior probability would place these hypotheses in this order. However, gray clouds are less likely on sunny days than rainy days, and about equally likely when it is rainy or there is a forest fire, so the likelihood favors rain or forest fire. The product of the prior and likelihood will favor rain, as it is both plausible and fits the observed data.

Learning structured representations

Bayes' rule can be applied to many different kinds of hypotheses, but some of the most impactful Bayesian models of cognition have focused on cases in which the hypotheses correspond to structured representations such as causal graphs or logical formulas. Bayes' rule is particularly useful in these cases, as it provides a way to explain how structured representations can be learned from data. Historically, learning structured representations have posed a challenge, leading advocates of structured representations to argue for strong innate constraints on learning (e.g., Chomsky, 1965) and advocate for learning to argue against structured representations (e.g., McClelland et al., 2010). Bayesian models offer a way to ask what structured representations should be (and can be) learned from data.

Bayesian inference over structured representations has proven particularly useful for explaining phenomena in categorization, causal learning, and language learning [see Causal Learning; Language Acquisition]. In each of these cases, different kinds of structured representations have been proposed as hypotheses considered by learners. In categorization, the hypotheses can be logical rules with prior probabilities depending on the complexity of those rules (Goodman et al., 2008). In causal learning, hypotheses correspond to different causal

structures (Griffiths & Tenenbaum, 2005). In learning language, hypotheses correspond to different grammars (Perfors et al., 2011). Being able to show which rules, structures, and grammars should be inferred from data provides a way to predict human behavior and engage with debates about learnability.

Exploring inductive biases

The role of the prior probability of hypotheses in Bayes' rule provides the opportunity to explore what people bring to the inductive problems they face in everyday life. Looking at Bayes' rule, the data d only appears in the likelihood $P(d|h)$, so the prior probability distribution $P(h)$ captures all of those factors other than the data that lead people to favor one hypothesis over another. In machine learning, these factors are referred to as the *inductive biases* of the learner. In cognitive science, those inductive biases might represent knowledge derived from other experiences, generic preferences for simpler hypotheses, or innate constraints that limit the set of hypotheses under consideration or favor specific hypotheses. By providing people with different data and seeing what hypotheses they select, Bayes' rule can be used to work backwards and make inferences about the prior probabilities of different hypotheses.

A simple example of inferring people's priors from their behavior is provided by analyzing everyday predictions (Griffiths & Tenenbaum, 2006). People are quite willing to make predictions about everyday events such as how much money a movie would make at the box office or how long it would take to bake a cake. By formulating this problem as one of Bayesian inference, it can be shown that different prior distributions result in different patterns of predictions. This makes it possible to compare people's predictions with the actual distributions of these everyday quantities. Similar approaches have been used to make inferences about prior distributions in other settings, including categorization (Goodman et al., 2008) and causal learning (Lu et al., 2008; for another approach to inferring priors, see Yeung & Griffiths, 2015).

Rich generative models

In its most general form, Bayes' rule can be applied to any probability distribution: the data are the aspects of the world that have been observed, and the hypotheses are everything else. Viewed in this way, all we need to apply Bayesian inference is a model that specifies the probabilities of the events that we want to make inferences from and about—a world model. One way to define such a model is to specify a procedure for generating observable events, called a generative model. This generative model can also include steps that are unobserved. These are called latent variables and can be inferred from the observable events.

For example, when predicting somebody else's behavior, it is useful to know what their goals are. This can be captured in a generative model in which people select a goal and then take actions to achieve that goal. The goal is a latent variable in this model, which can be inferred from people's actions (see Baker et al., 2009). Such a generative model can be extended to include many other factors relevant to interpreting people's behavior, such as people's preferences influencing the goals they choose [see Theory of Mind] (Jara-Ettinger et al., 2016).

Bayesian models of cognition have explored increasingly complex methods for defining increasingly rich generative models. Bayesian networks are a powerful tool for defining probability distributions that involve many variables and are very useful for clarifying the structure of generative models (Griffiths & Tenenbaum, 2006). Nonparametric Bayesian models make it possible to define generative models that do not assume the world is finite: for example, allowing for the possibility that there are an unbounded number of classes of objects (Sanborn et al., 2010). Probabilistic programs use ideas from programming languages to specify complex probability distributions that can incorporate recursion and facilitate probabilistic inference (Rule et al., 2020). Working with these more complex models typically requires using sophisticated methods for Bayesian inference developed in statistics and computer science such as Markov chain Monte Carlo or variational inference [see Markov Chain Monte Carlo].

Questions, controversies, and new developments

Bayesian models of cognition have been the target of several critiques focusing on the assumption of rationality, the links between Bayesian models and traditional mechanistic explanations, and the source of prior distributions (Jones & Love, 2011; Bowers & Davis, 2012; for responses, see Chater et al., 2011; Griffiths, Chater, et al., 2012). In addition, the relationship between Bayesian models of cognition and artificial neural networks is a point of discussion, as these models have seen increased popularity in artificial intelligence and machine learning.

Assumption of rationality

Bayesian models of cognition are based on the inferences that a rational agent should make from the available data. This assumption of rationality would seem to be at odds with Kahneman and Tversky's conclusion that people do not make probability judgments in accordance with Bayes' rule.

Advocates of Bayesian models of cognition have responded to this concern in several ways (in addition to Anderson, 1990). First, a strict correspondence to rationality in people's behavior is not necessary in order for Bayesian models to be useful. It is sufficient that people act in a way that is related to the rational solution to a degree that the solution still provides insight into their behavior. Even if there are systematic deviations from rationality, a Bayesian model can still be a useful first step in understanding some aspect of human cognition because it can help to make it clear what those systematic deviations are, just as knowing the prescriptions of probability theory was useful to Kahneman and Tversky in identifying people's biases and the corresponding heuristics that explained them.

Second, Kahneman and Tversky focused on probability judgment tasks that required people to make explicit assessments of the probability of various events based on other stated probabilities. Bayesian models of cognition have generally been applied to aspects of cognition other than explicit probability judgment, asking people more basic questions such as whether an object belongs to a category or whether a causal relationship exists. In these settings, Bayesian models have been empirically successful at predicting people's responses.

Even so, there are systematic deviations from rationality that emerge across a wide range of tasks: for example, people tend to *probability match*, producing responses with a probability that matches their posterior probability (e.g., Goodman et al., 2008).

Connections to cognitive mechanisms

A second line of criticism focuses on the fact that Bayesian models of cognition provide a different kind of explanation from that traditionally sought in psychological research. By focusing on the abstract problems that human minds face and their ideal solutions, Bayesian models of cognition operate at what Marr (1982) called the computational level. By contrast, psychologists have typically sought to explain human cognition at what Marr called the algorithmic level, focusing on the cognitive processes that underlie human behavior. This has led to confusion as well as concern: it seems implausible that people are explicitly computing Bayes' rule every time they face a problem of inductive inference, and if that is the case, what are the implications of Bayesian models of cognition for those who want to understand cognitive mechanisms?

Advocates of Bayesian models of cognition view them as a way of exploring what an agent should be doing when solving a problem and how that solution is affected by different assumptions about inductive biases. If a Bayesian model is found to be consistent with people's behavior, then it provides a clearer target for algorithmic-level models: whatever cognitive mechanisms are proposed, they need to be able to produce behavior consistent with that model. This kind of constraint has also led to modeling approaches that aim to bridge these two levels of analysis. Rational process models use algorithms from computer science and statistics that are known to approximate Bayesian inference as a source of hypotheses about the cognitive processes that people might engage in (Griffiths, Vul, & Sanborn, 2012). Resource rational analysis adds the expectation that these algorithms are deployed intelligently, supporting effective inferences while making efficient use of cognitive resources (Lieder & Griffiths, 2019). This approach can be used to explain some of the ways that people deviate from rationality. For example, the kind of probability matching people often demonstrate when making inductive inferences can be explained in terms of resource-rational use of a sampling algorithm (Vul et al., 2014).

Source of prior distributions

Prior distributions play an important role in Bayesian inference, and thus potentially provide a great deal of flexibility in trying to explain aspects of human cognition. Critics of Bayesian models of cognition have expressed concerns about how these prior distributions are identified.

One source of prior distributions is the environment in which human minds operate. In presenting the framework of rational analysis, Anderson (1990) focused on identifying reasonable prior distributions based on the environment. This approach is best exemplified by his subsequent work on memory (Anderson & Schooler, 1991). Griffiths and Tenenbaum (2006) took a similar approach to exploring people's predictions about everyday events, measuring the empirical distribution of a number of everyday quantities.

In other cases, Bayesian models are used to formulate theories about prior distributions that are tested through subsequent experiments. For example, Oaksford and Chater (1994) showed that people's errors in a logical reasoning task could be explained if people assumed it is unlikely that an arbitrary predicate will be true. They then conducted additional experiments to validate this assumption. Likewise, Goodman et al. (2008) evaluated several different prior distributions in their work explaining categorization as Bayesian inference over logical rules. In part, a Bayesian model of cognition represents a claim about what a relevant prior distribution might be. These claims are evaluated in the same way as other theories: by generating predictions and testing those predictions against data.

Finally, it is also possible to define Bayesian models in which the priors themselves are learned from the data. These hierarchical Bayesian models provide a way to understand "learning to learn," where more experience allows learners to make strong inferences from less data. This approach has been used to explain aspects of causal learning (Griffiths & Tenenbaum, 2009) and word learning (Kemp et al., 2007).

Relationship to artificial neural networks

As noted above, the growth and popularity of Bayesian models of cognition came in part from the development of more sophisticated tools for probabilistic reasoning in machine learning and artificial intelligence. More recently, these fields have come to be dominated by approaches based on artificial neural networks [see Large Language Models]. Artificial neural networks are also at the heart of another popular framework for modeling human cognition: connectionism. It is thus natural to ask how these two approaches relate.

The simple answer to this question is that there are a variety of ways in which artificial neural networks can implement Bayesian inference. Since Bayes' rule specifies the optimal solution to inductive problems, if a neural network is doing a good job of solving an inductive problem, it is likely to be approximating Bayesian inference. Marr's levels of analysis can also be appealed to in this setting, viewing Bayesian models of cognition as providing an abstract characterization of the ideal solution at the computational level and artificial neural networks as offering mechanisms by which those solutions might be carried out at the algorithmic or implementation levels.

Several interesting connections between Bayesian inference and artificial neural networks have been identified. Individual neural networks can be shown analytically to approximate simple forms of Bayesian inference (McClelland, 1998). Learning by gradient descent with various tweaks, such as weight decay or early stopping, can be interpreted in terms of a prior distribution on the weights of a network (MacKay, 1995), and those prior distributions can be tuned via metalearning (in which the initial weights of the network are trained to perform well across a distribution of tasks; Grant et al., 2018). Finally, neural networks can be explicitly trained to perform Bayesian inference in specific models, providing a way to amortize the costs of inference (Dasgupta & Gershman, 2021).

Broader connections

Bayesian models of cognition have close connections to ideal observer models in the study of perception. Many of the methods used in Bayesian models of cognition have also been used in probabilistic approaches to machine learning and artificial intelligence. More generally, the Bayesian approach has its roots in the broader literature on the philosophical foundations of statistics.

Further reading

- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243
- Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.
- Oaksford, M., & Chater, N. (Eds.). (1998). *Rational models of cognition*. Oxford University Press.
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004

References

- Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press. ↩
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x ↩
- Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005 ↩
- Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18327–18332. https://doi.org/10.1073/pnas.1306572110 ↩
- Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450 ↩
- Chater, N., Goodman, N., Griffiths, T. L., Kemp, C., Oaksford, M., & Tenenbaum, J. B. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral and Brain*

↩

- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

↩

- Dasgupta, I., & Gershman, S. J. (2021). Memory as a computational resource. *Trends in Cognitive Sciences, 25*(3), 240–251. https://doi.org/10.1016/j.tics.2020.12.008

↩

- Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology, 12*(1), 473–498. https://doi.org/10.1146/annurev.ps.12.020161.002353

↩

- Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review, 116*(4), 752–782. https://doi.org/10.1037/a0017196

↩

- Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32*(1), 108–154. https://doi.org/10.1080/03640210701802071

↩

- Grant, E., Finn, C., Levine, S., Darrell, T., & Griffiths, T. L. (2018, April 30–May 3). *Recasting gradient-based meta- learning as hierarchical Bayes* [Conference presentation]. ICLR 2018 Conference, Vancouver, Canada.

↩

- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*(4), 334–384. https://doi.org/10.1016/j.cogpsych.2005.05.004

↩

- Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17*(9), 767–773. https://doi.org/10.1111/j.1467-9280.2006.01780.x

↩

- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*(4), 661–716. https://doi.org/10.1037/a0017201

↩

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

↩

- Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science, 21*(4), 263–268. https://doi.org/10.1177/0963721412447619

↩

- Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naive utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences, 20*(8), 589–604. https://doi.org/10.1016/j.tics.2016.05.011

↩

- 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, 188–231. https://doi.org/10.1017/S0140525X10003134

↩

- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

↩

- Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*(3), 307–321. https://doi.org/10.1111/j.1467-7687.2007.00585.x

↩

- Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10*(7), 319–326. https://doi.org/10.1016/j.tics.2006.05.003

↩

- 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*, e1. https://doi.org/10.1017/S0140525X1900061X

↩

- Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*(4), 955–984. https://doi.org/10.1037/a0013256

↩

- MacKay, D. J. (1995). Bayesian neural networks and density networks. *Nuclear Instruments & Methods in Physics Research. Section A, Accelerators, Spectrometers, Detectors and Associated Equipment, 354*(1), 73–80. https://doi.org/10.1016/0168-9002(94)00931-7

↩

- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Freeman.

↩

- McClelland, J. L. (1998). Connectionist models and Bayesian inference. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 21–53). Oxford University Press.

↩

- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences, 14*(8), 348–356. https://doi.org/10.1016/j.tics.2010.06.002

↩

- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327–357. https://doi.org/10.1037/0033-295X.113.2.327

↩

- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608–631. https://doi.org/10.1037/0033-295X.101.4.608

↩

- Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118*(3), 306–338. https://doi.org/10.1016/j.cognition.2010.11.001

↩

- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68*(1), 29–46. https://doi.org/10.1037/h0024722

↩

- Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72*(3), 346–354. https://doi.org/10.1037/h0023653

↩

- Rule, J. S., Tenenbaum, J. B., & Piantadosi, S. T. (2020). The child as hacker. *Trends in Cognitive Sciences, 24*(11), 900–915. https://doi.org/10.1016/j.tics.2020.07.005

↩

- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review, 117*(4), 1144–1167. https://doi.org/10.1037/a0020511

↩

- Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review, 120*(2), 411–437. https://doi.org/10.1037/a0031912

↩

- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

↩

- Tenenbaum, J. B. (2000). Rules and similarity in concept learning. In S. A. Solla, T. K. Leen, K.-R. Müller (Eds.), *Advances in neural information processing systems 12* (pp. 59–65). MIT Press.

↩

- Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637. https://doi.org/10.1111/cogs.12101

↩

- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

↩

- Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology, 76*, 1–29. https://doi.org/10.1016/j.cogpsych.2014.11.001

↩