INQUIRY · LOGIC AND REASONING

Bayesian Reasoning in Everyday Inquiry

How Bayes' theorem describes learning from evidence, why people find it hard, and a simple method that makes Bayesian reasoning easy to apply in everyday decisions.

ModQu research article · 1,745 words · 17 references (APA 7th edition)

Every inquiry begins with some belief about how things are and then meets new evidence. A manager hears that a competitor is cutting prices; a doctor sees a positive test result; an analyst notices a sudden rise in complaints. The question in each case is the same: how much should this evidence change what I believe? **Bayesian reasoning** is the formal answer. It describes how a rational inquirer should combine what was already known with what has just been learned (Jaynes, 2003). It is also, as decades of research show, something people find surprisingly hard to do without help.

Bayes' theorem in plain terms

The theorem is named after Thomas Bayes, an eighteenth-century English minister whose essay on probability was published after his death by his friend Richard Price (Bayes, 1763). In modern notation, it states that the probability of a hypothesis *H* given evidence *E* is:

$$\mathrm{P}(\mathrm{H} \mid \mathrm{E}) = \mathrm{P}(\mathrm{E} \mid \mathrm{H}) \times \mathrm{P}(\mathrm{H}) \div \mathrm{P}(\mathrm{E})$$

Each term has an everyday meaning:

- **P(H), the prior:** how plausible the hypothesis was before the new evidence arrived. This is often a base rate, such as how common a disease is or how often a type of event happens.
- **P(E | H), the likelihood:** how probable the evidence would be if the hypothesis were true.
- **P(E):** how probable the evidence is overall, whether or not the hypothesis is true. It includes false alarms.
- **P(H | E), the posterior:** the updated belief after taking the evidence into account.

The key insight is that the strength of evidence depends not only on how well it fits the hypothesis, but also on how often the same evidence appears when the hypothesis is false, and on how plausible the hypothesis was to begin with.

Why intuition fails: neglecting base rates

In a famous study, Casscells et al. (1978) put a simple question to 60 students and staff at Harvard Medical School. If a test for a disease with a prevalence of 1 in 1,000 has a false positive rate of 5%, what is the chance that a person with a positive result actually has the disease? The most common answer, given by 27 of the 60, was 95%. Only 11 gave the

correct answer, which is about 2%. Out of every 1,000 people tested, about one has the disease, but about 50 healthy people also test positive. A positive result therefore raises the probability from 0.1% to about 2%, a large increase, but nowhere near 95%. Eddy (1982) described a similar error among physicians interpreting mammograms.

This error is known as **base-rate neglect**. Kahneman and Tversky (1973) showed that when people are given a vivid description of an individual, they tend to judge how likely it is that the person belongs to a category by how well the description fits, largely ignoring how common the category is. Tversky and Kahneman (1974) described this as one of several heuristics that make judgement fast but prone to systematic errors. Bar-Hillel (1980) found that people use base rates when they seem causally relevant to the case, but tend to ignore them when they appear to be mere background statistics. Koehler (1996) argued that the neglect is less general than early studies suggested.

A better format: natural frequencies

Gerd Gigerenzer and Ulrich Hoffrage argued that much of the difficulty lies in how the information is presented, not in human reasoning as such. Probabilities and percentages are a relatively recent invention; people are better at reasoning with counts of cases. In a series of experiments, Gigerenzer and Hoffrage (1995) gave participants problems such as the following:

> The probability of breast cancer is 1% for women aged 40 who take part in routine screening. If a woman has breast cancer, the probability that she will get a positive mammogram is 80%. If a woman does not have breast cancer, the probability that she will still get a positive mammogram is 9.6%. What is the probability that a woman with a positive mammogram actually has breast cancer?

The answer is about 7.8%. When the problem was stated as probabilities, about 16% of responses followed a correct Bayesian line of reasoning. When the same information was stated as **natural frequencies**, the figure rose to about 46%. Cosmides and Tooby (1996) reported similar gains, and a meta-analysis confirmed the effect across many studies (McDowell & Jacobs, 2017). Barbey and Sloman (2007) argue that the benefit comes from making the nested structure of the sets clear. Figure 1 shows the frequency version.

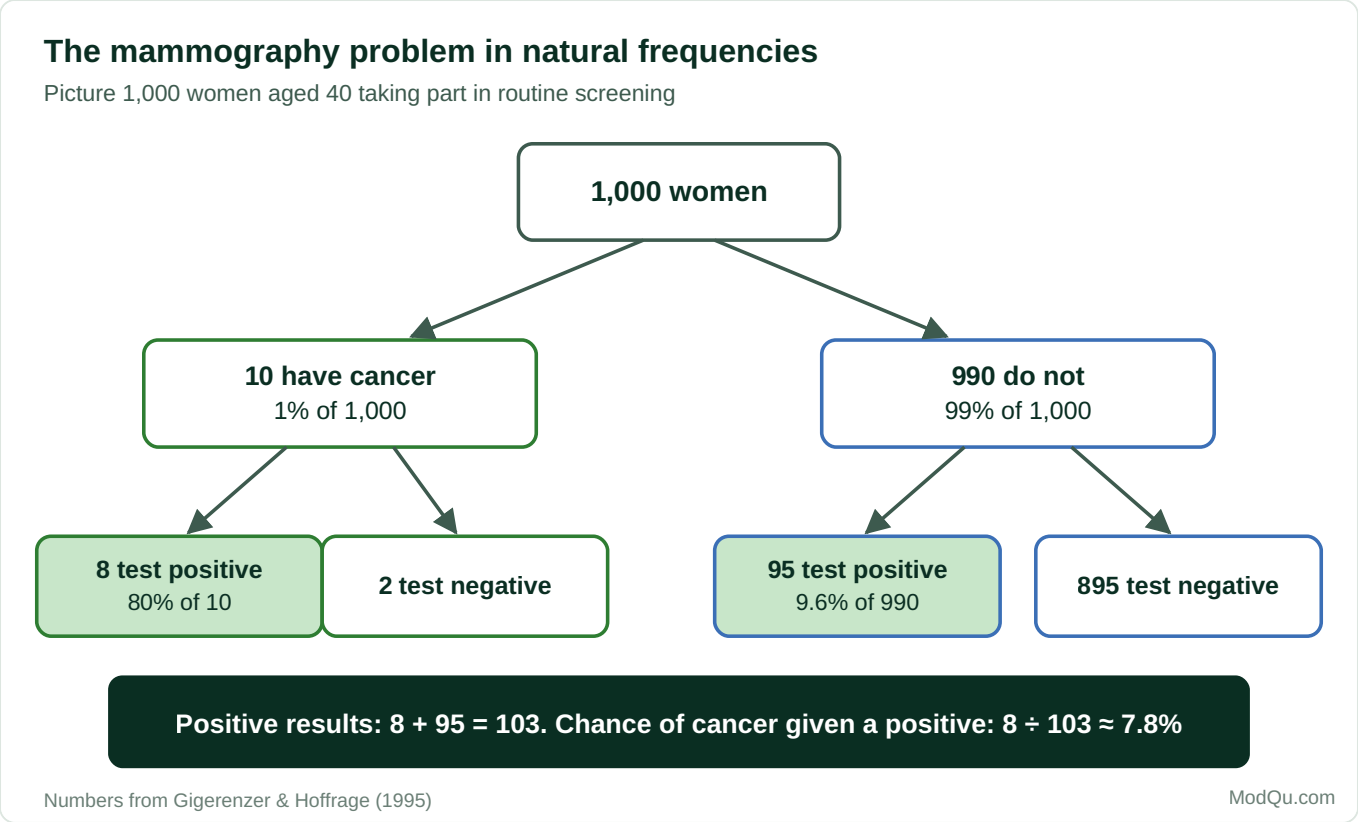


Figure 1. The mammography problem as a natural frequency tree. Of the 103 women who test positive, only 8 have cancer. Numbers from Gigerenzer and Hoffrage (1995); diagram by ModQu.

Laid out this way, the answer is visible: 8 true positives out of 103 positives in total. Hoffrage et al. (2000) showed that the same method helps physicians and legal professionals interpret diagnostic and forensic evidence, and recommended that statistical information be communicated as natural frequencies wherever possible (see also Gigerenzer et al., 2007). Brief training in the method produced lasting improvement (Sedlmeier & Gigerenzer, 2001).

A four-step method for everyday use

Bayesian reasoning does not require formulas in daily life. The following steps capture most of its value:

1. **Start with the base rate.** Before looking at the specific evidence, ask how common the thing is in general. How many leads usually convert? How often do complaints of this kind turn out to reflect a real product fault?
2. **Imagine a concrete group.** Picture 1,000 cases. How many have the feature you are interested in, and how many do not?
3. **Apply the evidence to both groups.** How many in each group would show the evidence? Be sure to count the false alarms as well as the hits.
4. **Compare.** Of all the cases that show the evidence, what share actually have the feature? That share is your updated belief.

A business example

Suppose 2% of visitors to a website eventually buy a product. A lead-scoring tool flags 90% of eventual buyers as "hot leads", but it also flags 10% of people who never buy. A sales team might assume that a hot lead is very likely to buy. Applying the four steps to 1,000 visitors:

- 20 will buy; the tool flags 18 of them.
- 980 will not buy; the tool flags 98 of them.
- Of the 116 hot leads, 18 will buy: about **16%**.

The tool is useful, since a hot lead is about eight times as likely to buy as an average visitor, but most hot leads will still not buy. A sales plan built on the intuitive reading of the tool would be badly wrong.

Updating step by step: the odds form

Bayes' theorem has a second form that is especially convenient when evidence arrives in pieces. Instead of probabilities, it uses **odds**, the ratio of the chance that something is true to the chance that it is false. In odds form, the theorem says:

$$\text{posterior odds} = \text{prior odds} \times \text{likelihood ratio}$$

The **likelihood ratio** is the probability of the evidence if the hypothesis is true, divided by its probability if the hypothesis is false. It measures how diagnostic the evidence is. A likelihood ratio of 1 means the evidence is equally likely either way and should not change belief at all; a ratio of 10 means the evidence is ten times more likely if the hypothesis is true.

Return to the Casscells problem. The prior odds of disease are 1 to 999. Assuming the test detects every true case, a positive result is 20 times more likely in someone with the disease (100%) than in someone without it (5%), so the likelihood ratio is 20. The posterior odds are 20 to 999, or about 2%, the same answer as before. Now suppose the person takes a second, independent test with the same properties, and it is also positive. The odds are multiplied by 20 again, giving 400 to 999, or about 29%. A third positive result would take the odds to 8,000 to 999, a probability of about 89%.

This example shows two features of Bayesian reasoning that matter in any inquiry. First, each piece of evidence moves belief by a fixed factor, so several moderately strong, independent pieces of evidence can together justify a confident conclusion that none could justify alone. Second, the pieces must genuinely be independent. Repeating the same test, asking the same source twice, or reading several news stories that all rely on one original report does not multiply the evidence, because the second observation adds little that the first did not.

Bayesian reasoning as an attitude of inquiry

Beyond calculation, Bayesian thinking expresses a general stance towards knowledge. Beliefs are held with degrees of confidence rather than as all-or-nothing positions. Each new piece of evidence moves those degrees up or down by an amount that depends on how diagnostic it is. Strong prior evidence is not overturned by a single surprising result, but repeated and independent evidence eventually changes even firm beliefs. Two people who start with different priors will, if they see enough of the same evidence, tend to converge (Edwards et al., 1963).

This stance has practical consequences for anyone running an inquiry:

- **State priors explicitly.** Writing down the expected answer before collecting data makes it easier to see how much the data changed things, and harder to claim afterwards that the result was obvious.
- **Ask how diagnostic the evidence is.** Evidence that would be equally likely under several hypotheses does not help distinguish between them, however striking it seems.
- **Watch for false alarms.** In any search for rare events, such as fraud, faults or unusual opinions, false positives can outnumber true ones even when the method is accurate.
- **Update incrementally.** Avoid both ignoring new evidence and overreacting to it. In early experiments people updated too little, a pattern called conservatism (Phillips & Edwards, 1966).

Limits

Bayesian reasoning depends on the quality of its inputs. A poorly chosen prior or a wrongly estimated false-positive rate will produce a confident but wrong conclusion, and in many everyday situations these quantities are known only roughly. Its value lies less in producing exact numbers than in forcing the inquirer to ask the right questions: how common is this, how often would I see this evidence if I were wrong, and how much should I really change my mind?

References

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211–233. https://doi.org/10.1016/0001-6918(80)90046-3

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*(3), 241–254. https://doi.org/10.1017/S0140525X07001653

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, *299*(18), 999–1001. https://doi.org/10.1056/NEJM197811022991808

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*(1), 1–73. https://doi.org/10.1016/0010-0277(95)00664-8

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.019

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*(3), 193–242. https://doi.org/10.1037/h0044139

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*(2), 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290*(5500), 2261–2262. https://doi.org/10.1126/science.290.5500.2261

Jaynes, E. T. (2003). *Probability theory: The logic of science* (G. L. Bretthorst, Ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511790423

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237–251. https://doi.org/10.1037/h0034747

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*(1), 1–17. https://doi.org/10.1017/S0140525X00041157

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin, 143*(12), 1273–1312. https://doi.org/10.1037/bul0000126

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

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*(3), 380–400. https://doi.org/10.1037/0096-3445.130.3.380

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124