Confirmation Bias

Jonathan D Nelson[^] and Craig R M McKenzie[+]

this entry appeared in pp. 167-171 of in M. Kattan (Ed.),

The Encyclopedia of Medical Decision Making.

London, UK: Sage

the full Encyclopedia can be purchased from Sage Publications,
http://www.sagepub.com/refbooksProdDesc.nav?prodId=Book231030;
it is ISBN 9781412953726

[^]nelson@mpib-berlin.mpg.de

Max Planck Institute for Human Development
Adaptive Behavior and Cognition Group
Lentzeallee 94
14195 Berlin, Germany

[+]cmckenzie@ucsd.edu

Rady School of Management and Department of Psychology
University of California, San Diego
La Jolla, CA 92093-0553

Confirmation bias is the tendency for people to search for or interpret information in a manner that favors their current beliefs. The goal of this entry is to communicate psychological research on confirmation bias as it relates to medical decision making. This will help medical professionals, patients, and policymakers to consider when it might pose a concern, and how to avoid it. The focus is on choosing a test for a simple case of medical diagnosis. The first section discusses how inference and information search ought to take place; the second section discusses confirmation bias and other possible errors; the final section discusses how to improve inference and information search.

How should inference and information acquisition proceed?

No choice of diagnostic tests can cause confirmation bias if the test results are assimilated in a statistically optimal manner. Therefore, this section first discusses how to incorporate test results in a statistically optimal (Bayesian) way. It then discusses various strategies to select informative tests.

Suppose that the base rate of a disease (d) in males is 10%, and that a test for this disease is given to males in routine exams. The test has 90% sensitivity (true positive rate), e.g. 90% of males who have the disease test positive. Expressed in probabilistic notation, $P(pos \mid d) = 90\%$. The test has 80% specificity, e.g. $P(neg \mid \sim d) = 80\%$ (20% false positive rate), meaning that 80% of males who do not have the disease correctly test negative. Suppose a male has a positive test in routine screening. What is the probability that he has the disease? By Bayes' theorem (see Figure 1A),

$$P(d \mid pos) = P(pos \mid d)\ P(d)\ /\ P(pos),$$

$$\text{where } P(pos) = P(pos \mid d)\ P(d) + P(pos \mid \sim d)\ P(\sim d).$$

Therefore,

$$P(d \mid pos) = (.90 \times .10)\ /\ (.90 \times .10\ +\ .20 \times .90) = .09\ /\ .27 = 1/3.$$

Alternately (see Figure 1B and Figure 1C), it is possible to count the number of men with the disease and a positive test, and who test positive without having the disease:

$$P(d \mid pos) = \text{num}(d \,\&\, pos) \,/\, \text{num}(pos) = 9 \,/\, (9 + 18) = 1/3, \text{ where}$$
$$\text{num}(pos) = \text{num}(pos \,\&\, d) + \text{num}(pos \,\&\, {\sim}d).$$

But how should a diagnostic test be chosen in the first place? The fundamental difficulty is that which test is most useful depends on the particular outcome obtained, and the outcome cannot be known in advance. For instance, the presence of a particular gene might definitively predict a disease, but that gene might occur with only one in a million probability. Another test might never definitively predict the disease, but might always offer a high degree of certainty about whether the disease is present or not.

Optimal experimental design ideas provide a reasonable framework for calculating which test, on balance, will be most useful. All of these ideas are within the realm of Savage's Bayesian decision theory, which defines the subjective expected usefulness (utility) of a test, before that test is conducted, as the average usefulness of all possible test results, weighting each result according to its probability.

In the case of a test *T* that can either be positive (*pos*) or negative (*neg*), the test's expected utility (eu) would be calculated as follows:

$$eu(T) = P(pos) * u(pos) \;+\; P(neg) * u(neg),$$

where u corresponds to utility. Various optimal experimental design ideas quantify, in different ways, the usefulness of particular test outcomes. Suppose one wishes to use improvement in probability of correct diagnosis to quantify the usefulness of possible diagnostic tests. (This equates to minimizing error.) The probability gain (pg) of a test, with respect to determining whether or not a patient has disease *d*, is calculated as follows:

$$eu_{pg}(T) = P(pos) * [\ \max(\ P(d \mid pos), P({\sim}d \mid pos)\ ) - \max(\ P(d), P({\sim}d)\ )\ ]$$
$$+ P(neg) * [\ \max(\ P(d \mid neg), P({\sim}d \mid neg)\ ) - \max(\ P(d), P({\sim}d)\ )\ ].$$

Suppose the goal is to learn whether or not a patient has a disease which occurs in 10% of patients. Test 1 has 95% sensitivity and 85% specificity. Test 2 has 85% sensitivity and 95% specificity. Which test maximizes probability gain? Test 1 has probability gain zero, although Test 2 has probability gain 0.04. Although Test 1 has high sensitivity, its low specificity is problematic as the base rate of the disease is only 10%. Test 1 does not change the diagnosis of any patient, because irrespective of whether it is positive or negative, the patient most likely does not have the disease. Test 2's much higher specificity, however, reduces false positives enough that a majority of people who test positive actually have the disease.

It can be helpful, as an exercise, to consider possible tests' probability gain before ordering a test, in situations where the relevant environmental probabilities are known. In real medical diagnosis additional factors, such as a test's cost and its potential to harm the patient, should also be taken into account.

Confirmation bias and other errors

Do people typically reason following Bayes' theorem? Do physicians intuitively select useful tests for medical diagnosis? If human cognition and behavior are suboptimal, do they reflect confirmation bias?

From early research on Bayesian reasoning through the present, there has been evidence that people are either too conservative or too aggressive in updating their beliefs. Some research suggests that people make too much use of base rates (the proportion of people with a disease), as opposed to likelihood information (a test result). Other research suggests that people make too little use of base rates, relying on likelihood information too much.

Do these errors lead to systematically overweighting one's working hypothesis (e.g. the most probable disease)? Note that test results can either increase or decrease the probability of a particular disease. Because of this, neither being too conservative nor being too aggressive in updating beliefs in response to test results would consistently give a bias to confirm one's working hypothesis. Thus, while there is plenty of evidence that people (including physicians) sometimes update too much and sometimes too little, that does not necessarily imply confirmation bias.

If people have personal experience with environmental probabilities, their inferences are often quite accurate. In routine diagnostic and treatment scenarios, in which individual practitioners have previously experienced dozens, hundreds, or even thousands of similar cases and have obtained feedback on the patients' outcomes, physicians' intuitions may be well-calibrated to underlying probabilities. Little if any confirmation bias would be expected in these situations. In situations in which relevant data are available but practitioners do not have much personal experience, for instance because rare diseases are involved, intuitions may not as closely approximate Bayes' theorem.

Confirmation bias in inference. Apart from the general difficulty in probabilistic reasoning, how might people fall victim to confirmation bias per se? Below, several situations are described that might lead to confirmation bias.

1. If people obtain useless information, but think it is supports their working hypothesis, that could lead to confirmation bias. Suppose a physician asks a patient about the presence of a symptom which, if present, would support a particular disease diagnosis. Suppose the patient tends to answer "yes" in cases where the question is unclear, so as to cooperate. If the physician does not take the patient's bias to answer "yes" into account when interpreting the answer to the question, the physician could be led, on average, to be excessively confident in their diagnosis.

2. Sometimes a test's sensitivity (its true positive rate) is conflated with its positive predictive value (the probability of the disease given a positive result). In situations where the sensitivity is high, but specificity is low or the base rate of the disease is very low, this error can cause confirmation bias. For instance, among people from low-risk populations, a substantial proportion of people with positive HIV test results do not have HIV. Some counselors, however, have wrongly assumed that a positive test means a person has HIV.

3. There are many situations in which people want to reach certain conclusions or maintain certain beliefs, and they are quite good at doing so. Imagine that a physician has diagnosed a patient with a serious illness and started the patient on a series of treatments with serious side effects. The physician might be more likely than, say, an impartial second physician, to discount new evidence indicating that the original diagnosis was wrong and that the patient had needlessly been subjected to harmful treatments.

4. Finally, people sometimes interpret ambiguous evidence in ways that give the benefit of the doubt to their favored hypothesis. This is not necessarily a flaw in inference. If one's current beliefs are based on a great deal of information, then a bit of new information (especially if from an unreliable source) should not change beliefs drastically. Whether a physician interprets a patient's failure to return a smile from across the room as indicating the patient didn't see them or as a snub will likely be influenced by whether the patient has previously been friendly or socially distant. Similarly, suppose an unknown researcher emails their discovery that AIDS is caused by nefarious extraterrestrials. Given the outlandish nature of the claim, and the unknown status of the 'researcher', it would be wise to demand a lot of corroborating evidence before updating beliefs about causation of AIDS at all, given this report. The overriding issue is that one's degree of belief, and amount of change of belief, should correspond to the objective value of the evidence.

Information acquisition and confirmation bias. Do people use statistically justifiable strategies for evidence acquisition, for instance when requesting a test or asking a patient a question? Are people prone to confirmation bias or other errors?

Psychological experiments suggest that people are very sensitive to tests' usefulness when deciding which test to order. Any testing strategy not solely concerned with usefulness will be inefficient. However, if test results are evaluated in a Bayesian way, then although some information-acquisition strategies are more efficient than others, none will lead to confirmation bias. Thus, improving probabilistic inference is a first step towards guarding against confirmation bias.

Positivity and extremity are additional factors that may contribute to people's choices of tests. Positivity is the tendency to request tests that are expected to result in a positive result, or a “yes” answer to a question, given that the working hypothesis is true. Extremity is a preference for tests whose outcomes are very likely or very unlikely under the working hypothesis relative to the alternate hypothesis. The evidence substantiating people's use of these particular strategies is somewhat murky. However, use of these testing strategies, together with particular biased inference strategies, could lead to confirmation bias.

Improving inference and information acquisition

Improving inference. The means by which probabilistic information is presented are important, and evidence suggests that either personal experience or appropriate training can help people meaningfully learn particular probabilities. The literature suggests several strategies to improve Bayesian inference:

1. Present information in a meaningful way. Figure 1, panels B and C, illustrates two means of presenting equivalent information, in which the information is presented in terms of the *natural frequencies* of people with (and without) the disease who have a positive or negative test. These formats better facilitate Bayesian reasoning than does the standard probability format (Figure 1A). Simulating personal experience and providing feedback may be even more effective.

2. Teach Bayesian inference. Although people do not intuitively do very well with standard probability format problems, people can be trained to do better, especially when the training helps people utilize natural frequency formats for representing the probabilistic information.

3. Obtain feedback. Feedback is critical for learning environmental probabilities, such as base rates of diseases, and distribution of test outcomes for people with and without various diseases. Feedback is also critical for learning when those probabilities change, for instance because of an outbreak of a rare disease. Both individual practitioners and

policymakers could think about how to ensure that feedback can be obtained, and patients and citizens should demand that they do so.

Improving information acquisition. People are not adept at maximizing either probability gain or individually specified utilities when information is presented in the standard probability format. Taking care to ensure that known statistical information is meaningful may be the single most important way to improve practitioners' capacity for good inference and information acquisition in medical decision making. Use of personal experience and feedback to convey probabilistic information in simulated environments can also facilitate Bayesian performance.

Beyond confirmation bias

While confirmation bias in inference and information acquisition may exist, it should be seen in the broader context of statistical illiteracy and misaligned incentives. Those problems may be the root of what can appear to be confirmation bias, rather than any inherent cognitive limitations that people have. For instance, the desire to make a patient feel that they are being treated well, and to guard against possibility of litigation, might lead to ordering a medically unnecessary (and potentially harmful) CT scan following mild head trauma. At the level of basic research, the source of funding can influence the conclusions that are reached. From a policy standpoint, the goal should be to make individual and institutional incentives match public health objectives as closely as possible.

Related entries

Also see entries on Bayes's Theorem, Biases in Human Prediction, Cognitive Psychology and Processes, Conditional Probability, Deliberation and Choice Processes, Errors in Reasoning or Judgment, Evidence Synthesis, Expected Value of Information, Expected Utility Theory, Hypothesis Testing, Probability , Probability Errors, Subjective Expected Utility Theory

Further reading

Baron, J. (1985). *Rationality and intelligence*. Cambridge, England: Cambridge University Press.

Baron, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: I. Priors, error costs, and test accuracy. *Organizational Behavior and Human Decision Processe*s, 41, 259–279.

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, 42, 88–110.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, l02(4), 684-704.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (in press). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*.

Kahneman, D, & Tversky, A (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430-454.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, 32, 385-418.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information. *Psychological Review*, 94, 211–228.

McKenzie, C. R. M. (2004). Hypothesis testing and evaluation. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 200-219). Oxford: Blackwell

McKenzie, C. R. M. (2006). Increased sensitivity to differentially diagnostic answers using familiar materials: Implications for confirmation bias. *Memory and Cognition*, 34, 577-588.

Nelson, JD (2005). Finding useful questions: on Bayesian diagnosticity, probability, impact and information gain. *Psychological Review*, 112(4), 979-999.

Nelson, JD (2008). Towards a rational theory of human information acquisition. In Oaksford, M & Chater, N (Eds.), *The probabilistic mind: Prospects for rational models of cognition* (pp. 143-163). Oxford: Oxford University Press.

Nickerson, R. S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin & Review*, 10, 289–318.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Sedlmeier, P, & Gigerenzer, G (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, 130, 380–400.

Figure 1.

Different formats for presenting probabilistic information.

The means in which probabilistic information is presented has a large impact on how meaningful it is to people. The standard probability format (panel A) is complicated for people to work with, although people can be trained to do so. Both the frequency tree (panel B) and frequency grid diagram (panel C) provide more meaningful representations of the information. '*d*' denotes the disease; ~*d* absence of the disease; '*p*' denotes a positive test. In panel C, shaded cells denote presence of the disease.

A. Standard format   B. Frequency tree   C. Frequency grid diagram

$P(d) = 10\%$

$P(\sim d) = 90\%$

$P(p \mid d) = 90\%$

$P(p \mid \sim d) = 20\%$