Alan Jessop

Let the Evidence Speak

Using Bayesian Thinking in Law, Medicine, Ecology and Other Areas

Springer

Let the Evidence Speak

Alan Jessop

Let the Evidence Speak

Using Bayesian Thinking in Law, Medicine, Ecology and Other Areas

Springer

Alan Jessop
Durham University Business School
Durham University
Durham, United Kingdom

ISBN 978-3-319-71391-5 ISBN 978-3-319-71392-2 (eBook)
https://doi.org/10.1007/978-3-319-71392-2

Library of Congress Control Number: 2017960423

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Chapter 1
Introduction

Bayes is all the rage!

You'll read this, or something like it, from time to time. This example is from the opening sentence of a book by Luc Bovens and Stephan Hartmann [1]. In full the sentence reads "Bayes is all the rage in philosophy". You may not be much interested in philosophy, but that's just one of the places where Bayes' Rule makes an appearance. There are others, not only in books written for academics.

For instance, Angela Saini wrote a piece in *The Guardian* about the use of Bayes' Rule in a court of law under the headline *The formula for justice*. She writes that Bayes' Rule was

> Invented by an 18th-century English mathematician, Thomas Bayes, this calculates the odds of one event happening given the odds of other related events [2].

And here's Tim Harford writing in a piece called *How to make good guesses* that when we have two pieces of information

> Logically, one should combine the two pieces of information ... There is a mathematical rule for doing this perfectly (it's called Bayes' rule) [3].

Finally, describing the application of Big Data to online advertising, Cathy O'Neil writes

> The data scientists start off with a Bayesian approach, which in statistics is pretty close to plain vanilla. The point of Bayesian analysis is to rank the variables with the most impact on the desired outcome. Search advertising, TV, billboards, and other promotions would each be measured as a function of their effectiveness per dollar. Each develops a different probability, which is expressed as a value, or a weight [4].

All of which is fine, but just what is Bayes' Rule?

It's been with us for about 300 years and every so often enjoys some popular prominence. Articles and books such as those above are not written for a technical audience. They may have whetted your appetite—sounds good but how does it work? You want to know more but you don't want a statistics text. This book is for you.

~ • ~

A. Jessop, *Let the Evidence Speak*, https://doi.org/10.1007/978-3-319-71392-2_1

Decisions should be made based on evidence.

A doctor may use a test such as a blood test or MRI scan when diagnosing what illness a patient might have. Experimental results show the diagnostic power of the test, its ability to provide evidence for the doctor who then must decide what to tell the patient. Some tests are very informative, some less so.

Police investigating a crime and juries trying an accused have to use evidence too. This is of variable quality, ranging from DNA matching (pretty good) to eyewitness identification (not so good), and yet decisions of guilt and innocence are made.

What is common is that someone—a doctor, a juror—makes a decision based on what they believe to be true and this belief is based, in part at least, on evidence. Bayes' Rule provides the necessary link between the evidence and what to believe based on that evidence. The better the evidence the stronger is the belief that the patient is ill or the accused is guilty. The strength of the evidence, and so of our belief, is described using probability.

Bayes' Rule takes us that far. What to do, what action to take, is your decision. You will have a measure of your justified belief to help you.

~ • ~

There are many fine books and research papers you could read about Bayes' Rule and its applications but they tend to be written in the formal language of mathematics. They may assume that you know more than you do. We all hit a barrier at some point.

The motivation for this book is that it is possible and useful to describe for any reader, whatever their (which is to say, your) background, what Bayes' Rule is and how it works. Bayes' Rule is usually presented as a formula. For many this is not an encouragement to read further. But a great many real applications, not just text book illustrations, need no more than simple arithmetic. To show the calculations I have used a table—the Bayes Grid. All you need to know is how to add, multiply and divide. Think of a very simple spreadsheet.

Bayes' Rule is also a way of thinking, of forcing you to answer questions

What do I know about this problem?
What evidence do I have and how good is it?
What alternative explanations might account for the evidence?

Asking the right questions is always a good idea, even if you sometimes need help with the maths.

~ • ~

This book is organised in three parts. The first sets out the basic relation between evidence and the alternative accounts for that evidence. The second extends the basic framework to bring in other information you may have. This may be based on data or on judgement. You need to know how to deal with the issues raised by this useful extension.

Up to this point the calculations have been easy. The emphasis has been on thinking about how to use evidence, asking the right questions. Many applications are not so straightforward either because the maths is more difficult or because the decision problem is, by its nature, more complex. The chapters in this third part cover both.

In summary, the three parts are

Likelihood The key to it all. The necessary description of the relation between the evidence you have, what might account for it, and what you are justified in believing.

Base rate Base rates enable you to use contextual or judgemental information as well as evidence. Bayes' Rule becomes a means of learning, modifying your starting belief in light of new evidence. Bringing in judgement is quite natural, but care is needed. We may not be as good as we think at expressing our judgement in the form of probabilities.

Application Many of the applications of Bayes' Rule use mathematical models which are daunting for non-mathematicians. However, the underlying principles are the same as in the earlier parts. These final chapters show how some Bayes' thinking helps structure problems and some of the modelling issues which follow.

The closing two chapters, on law and the psychology of reasoning, do not involve heavy maths. They further emphasise the benefits of using Bayes' Rule as a way of thinking.

Notes and references are there to provide hints and reading should you wish to go further. Some of the references may not be for you but none are needed for you to read this book and, I hope, benefit from it.

~~~ • ~~~

References

1. Bovens L, Hartmann S (2003) Bayesian epistemology. Clarendon, Oxford, p v
2. Saini A (2011) The formula for justice. Guardian Magazine, 2 Oct 2011, p 12
3. Harford T (2016) How to make good guesses. Financial Times Weekend Magazine, 27/28 Feb, p 45
4. O'Neill C (2016) Weapons of math destruction: how big data increases inequality and threatens democracy. Allen Lane, London, p 74
~~~

Part I
Likelihood

Chapter 2
Whose Car?

Evidence provided to the police is often imperfect. Even witnesses who are quite sure of what they saw or heard may be wrong. How should their evidence be evaluated?

Let us suppose.

There has been a robbery at a jewellers in Stockholm. Inspector Larsson and his squad have narrowed the suspects down to just two, Jan and Stig. Larsson is sure that one of them is guilty but there is no conclusive evidence to decide between them. Circumstantial evidence puts them both near the crime scene at about the right time, but that's not enough. Then a witness, Ingrid, comes forward. She was walking in the street where the robbery occurred and saw a car driving very fast away from the jewellers. She only got a very brief look at the car and all she could say was that it was blue. Stig's car is blue and Jan's car is green. But thanks to the creative efforts of car manufacturers it's not quite as clear cut as that. A list of car paint codes shows Albi Blue, Storm Blue, Mountain Blue, Odyssey Blue, Spray Blue. . .. There are twenty-six shades with blue in the name, and then there's Aqua and others. There are only eleven which are green, including Nordic Green. Larsson smiles. Probably all that can be said is that Stig's car is more blue-ish and Jan's car more green-ish.

So, how can Ingrid's evidence be evaluated?

Colour is defined by wavelength measured in nanometres (nm). One nanometre is a billionth of a metre. The visible spectrum is from about 400nm to about 700nm. Shorter wavelengths, less than 400nm, are called gamma rays and X rays. Longer wavelengths are used for microwave and radio. In the visible part of the spectrum ranges of wavelength are given names which are the colours with which we are all familiar (think rainbow)[1]:

[1]These values are from the NASA website. https://science-edu.larc.nasa.gov/EDDOCS/Wavelengths_for_Colors.html. Accessed 14 September 2017.

A. Jessop, *Let the Evidence Speak*, https://doi.org/10.1007/978-3-319-71392-2_2

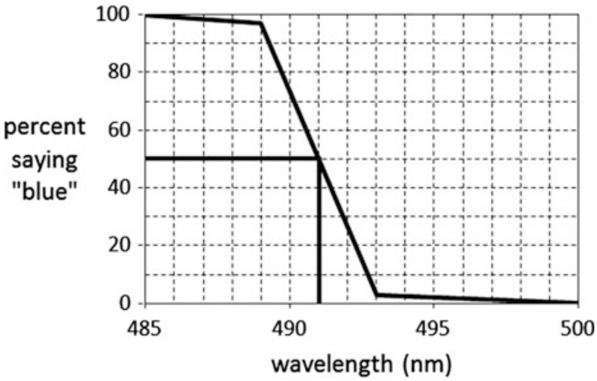
violet	about 400
indigo	about 445
blue	about 475
green	about 510
yellow	about 570
orange	about 590
red	about 650

The way we map this infinite variety into a small and manageable number of colours is called *categorical perception*. The same idea can be applied to sounds and other stimuli. There is inevitably some uncertainty, some vagueness, about where to draw the line. We may be inconsistent, applying different labels to the same stimulus on different occasions. Different people will apply different labels. When Ingrid said "blue" what did she mean?

While browsing the web Larsson's trusted colleague, Mankell, comes across a study [1] which may help. In it a number of people were shown colour cards for a short time and then asked to say whether what they saw was blue or green. The results are shown as *identification functions*. For any colour shown by its wavelength on the horizontal axis we can see the percentage of times it was identified either as "blue" or as "green".

Figure 2.1 shows the function for the label "blue". As the wavelength increases fewer people say "blue". For a colour with wavelength 491nm expect that fifty percent of people would call it "blue".

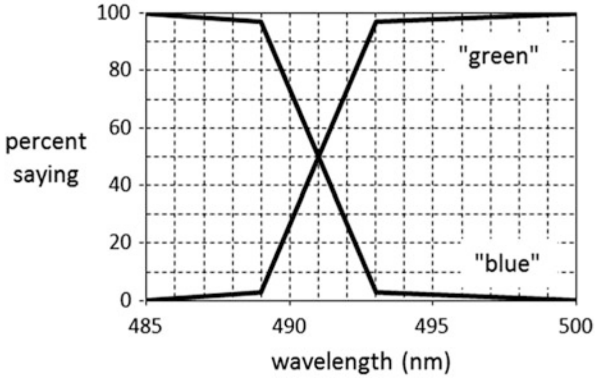
Fig. 2.1 Identification function for "blue"[2]



[2]This and other graphs in this chapter use information from Fig. 2a of Bornstein and Korda [1]: 212.

Figure 2.2 shows the identification functions for both “blue” and “green”. The two lines are complementary, mirror images. For any wavelength the two percentages sum to one hundred since only those two alternative names—“blue” and “green”—were used in the experiment.

Fig. 2.2 Identification functions for “blue” and “green”

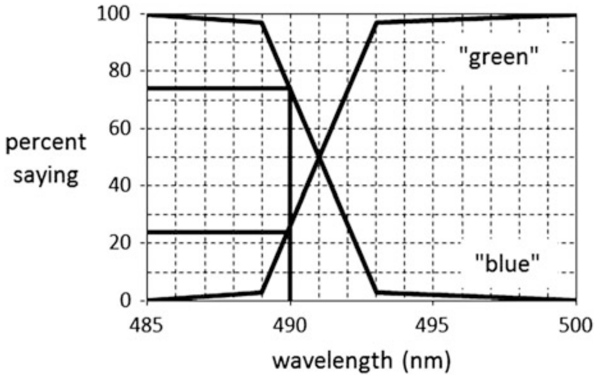


This graph might be helpful in assessing just what weight should be given to Ingrid’s evidence. The actual colours of the two cars are known, the wavelengths have been measured by the crime scene investigation team, and so the percentage of times that each car would be expected to be called “blue” or “green” can be read from the graph.

There is no reason to think that Ingrid differs in any relevant way from the subjects of the original experiment. This graph will do as a basis for assessing how useful her evidence is in discriminating between Stig’s car and Jan’s car.

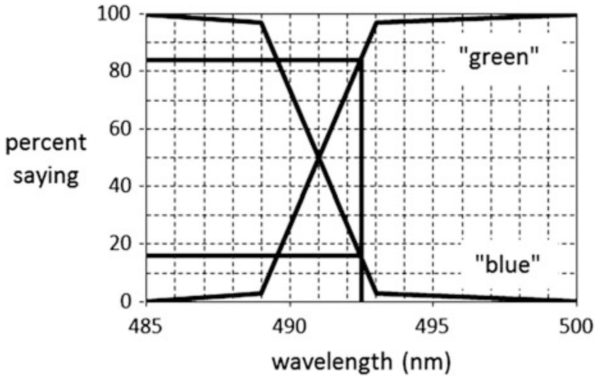
We can mark the actual colour of Stig’s car (490nm). Our best estimate is that seeing Stig’s car seventy-four percent of people will call it “blue” and twenty-six percent will say “green” (Fig. 2.3).

Fig. 2.3 Stig’s car



For Jan's car (colour is 492.5nm) Fig. 2.4 shows that only sixteen percent of people would call it "blue" and eighty-four percent "green".

Fig. 2.4 Jan's car



Because they assume that Ingrid is typical of the subjects tested in the experiment it seems clear to Larsson and Mankell that these results may be used to describe the accuracy of her assessment: when she says "blue" the odds that the car is Stig's car rather than Jan's car are 74:16. To put it another way, the chance that the car was Stig's car is eighty-two percent.[3] The chance that it was Jan's car is eighteen percent. Larsson and Mankell conclude that based on Ingrid's evidence the probability that it was Stig's car that she saw is about eighty percent. Not bad.

~ • ~

The two policemen did what seemed obvious. Were they right?

Yes, they were.

Although they didn't know it, they had used Bayes' Rule. It was the intuitively obvious thing to do.

~ • ~

This is fine so far as it goes. We have a good idea of the strength of this evidence but how should it be used together with what else we know, or suspect? How to combine different sources of evidence?

The natural language of the courtroom is not numeric and neither judges nor jury members (nor many of us) are used to this form of reasoning. But courts do hear evidence from expert witnesses, much of which is based on quantitative analysis, and juries do make decisions.

[3]Odds and probabilities both express uncertainty. The conversion from one to the other is easy. 74/(74 + 16) = 82% and 16/(74 + 16) = 18%.

The usefulness and, which is different, the admissibility of arguments about the possible statistical basis of evidence and the justifiable evaluation of that evidence is subject to quite a bit of argument. Justification is important.

∼ • ∼

Stig, Jan, Ingrid and the rest are fiction. But the possible use of Bayes' Rule to assess evidence and, more problematically, to present the analysis in court, are not.

In 1996, Denis Adams was accused of rape. The victim said her attacker was in his twenties. Adams was 37. The victim was unable to pick Adams in an identity parade. Adams' alibi was that he had spent the night in question with his girlfriend. But DNA evidence at the scene was a good match with Adams' DNA, though there was some dispute about the correct match probability. The match probability is the probability that someone picked at random would match the DNA profile found at the crime scene. The prosecutor said 1 in 200 million but the defence argued for a much lower figure of 1 in 20 million or perhaps even 1 in 2 million. All other evidence supported Adams' claim of innocence. What was the jury to make of all this? Professor Peter Donnelly of Oxford University was permitted by the court to show the jury how Bayes' Rule could be used to evaluate the evidence presented to them. Adams was convicted. He appealed against his conviction.

The Court of Appeal noted that at the original trial no advice was given to the jury about what they should do if they chose not to use Bayes' Rule. A retrial was ordered.

At the retrial the court asked that Professor Donelly and other statisticians on the prosecution's side cooperate to provide a guide to help the jury. Although Donnelly had reservations a method was agreed. Jury members were required to complete questionnaires to illustrate how Bayes' Rule could be used (Professor Donelly gives a brief account of all this in the Royal Statistical Society's journal *Significance* [2]).

Adams was convicted again. He appealed again. The appeal was not upheld. But the Court of Appeal made observations critical of the use of Bayes' Rule. For example

> Jurors evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them.

> Individual jurors might differ greatly not only according to how cogent they found a particular piece of evidence (which would be a matter for discussion and debate between the jury as a whole), but also on the question of what percentage figure for probability should be placed on that evidence.

> Different jurors might well wish to select different numerical figures even when they were broadly agreed on the weight of the evidence in question.

The general drift was that the use of Bayes' Rule, or any mathematical model, would illegitimately undermine "an area peculiarly and exclusively within the province of the jury, namely the way in which they evaluate the relationship between one piece of evidence and another". In other words, the mental processes of twelve members of the public can be trusted to make some quite complicated judgements. That is sometimes inevitable, of course, but where a little calculation might help it seems harsh to deny the jurors the use of it.

The court decided against the use of Bayes' Rule in assessing DNA evidence. But, even if jurors had some appreciation of probabilities of events familiar to them or which they might easily imagine—tossing a coin, choosing a card—they (and we) are unlikely to have an appreciation of what a probability of 1 in 200 million, or 0.000000005, means. In the trial this very low rate implied that the number of people in the UK matching Adam's DNA profile was small, just one or two. They might be children or older folk and so not suspects. Or they might be related to Adams. He did, in fact, have a half-brother whose DNA profile was not known.

If the match rate was as high as 1 in 2 million then there might be about thirty or so people who would match: a different picture.

Guidance along these lines has subsequently been issued to judges so that they might help jurors, and themselves, get a grip on these very low probabilities.

~ • ~

This isn't a problem which will go away. In 2010 a convicted killer, "T", took his case to the Court of Appeal. One of the pieces of evidence against him involved his Nike trainers and the likelihood that shoe prints at the scene of the crime matched his shoes. The judge believed that the expert witness had incorrectly calculated the match probability. The case was quashed.

Professor Norman Fenton of Queen Mary, University of London, pointed out that the conclusion was not well justified and, surprise, proposed that Bayes' Rule would help.[4] He, with others, built a model based on the rule to help in judicial decision making. Their model allows the combination of evidence which is based on data and evidence which relies on judgement.[5]

We'll return to this problem of using Bayes thinking in court in Chap. 16.

~ • ~

The same ideas should help doctors to make diagnoses. You can easily imagine other applications.

And just think where all that big data analysis might lead. Simply having more data is not enough, we still need to have a way of reaching justifiable decisions.

It should be clear that numerical analyses do not replace judgement. But where the application of a little mathematics can help that judgement it would be wilful to reject it out of hand. There is pretty good evidence that we human beings aren't too smart when it comes to dealing with numerical information unaided by some calculation. Just what to do with the answers is where judgement comes to the fore.

~ • ~

[4]Some of the issues raised are discussed in [3].

[5]More details from their company, Agena. Bayesian Network and Simulation Software for Risk Analysis and Decision Support. http://www.agenarisk.com/. Accessed 14 September 2017.

The concerns of the courtroom were a long way from the mind of the Reverend Thomas Bayes when he devised his rule for reasoning from evidence to belief. But just what is Bayes' Rule? And who was Bayes?

~~~ ••• ~~~

References

1. Bornstein MH, Korda NO (1984) Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. Psychol Res 46(3):207–222
2. Donelly P (2005) Appealing statistics. Significance 2(1):46–48
3. Fenton NE, Neil M, Hsu A (2014) Calculating and understanding the value of any type of match evidence when there are potential testing errors. Artif Intell Law 22(1):1–28
~~~

Chapter 3
Bayes' Rule

In the previous chapter the two policemen had evidence that was not precise. The evidence they had was an eyewitness report that a car was "blue" and they wanted to see how strongly this pointed to Stig's car, which was blue, or to Jan's car, which was green. This is typical of many problems we face: we have to decide between a number of alternatives (which car?) based on some imprecise evidence ("blue"). To judge just how useful the evidence is we need to know the relation between each of the alternatives and the evidence. This goes by many names: accuracy, track record, diagnostic power and so on.

The idea is simple enough. If I don't know how much you know about the Nigerian economy why should I ask you about it? Or, if you give your opinion anyway, how seriously should I take what you say? I should find out about your track record. How good a pundit are you? Often we don't, of course. What do you know of your doctor's track record? Dealing with this sort of judgemental evidence raises all sorts of issues about credibility, bias and much else.

Fortunately, for a great many problems we have data on which to base an evaluation of the relation between what we have as evidence and the alternative causes or explanations of that evidence. Larsson and his team show how we might use evidence. It is a straightforward case which we can put into a more formal, though still simple, framework which will be useful for other problems.

Remember that the paint used by the makers of Stig's car would be called "blue" by seventy-four percent of people and "green" by the rest. Only sixteen percent of people would call Jan's car "blue". Table 3.1 shows this information.

A. Jessop, *Let the Evidence Speak*, https://doi.org/10.1007/978-3-319-71392-2_3

alternatives: car colour	*evidence*: witness statement "blue"	"green"	sum
Stig's car	74	26	100%
Jan's car	16	84	100%

Table 3.1 Colour identification depends on car

Each row of the table shows how likely is the evidence—"blue" or "green"—depending on which car was seen. The sum of each row is one hundred percent. This is because only these two values of evidence are thought to be possible. We have ruled out the possibility that a witness will say that either car was red, for instance. So, a witness must say either that the car was blue or that it was green: there are no other alternatives, no other colours are considered. These two distinct values cover all possibilities. That is why the sum is one hundred percent.

This is a table presenting some experimental data.

The witness, Ingrid, was not part of the study described in the research but we have no reason to believe that she is different from the study participants in the way she identifies colours. It is reasonable to say that the likelihood, the probability, that she would have called Stig's car blue is seventy-four percent.

Just pause for a minute. The shift in perspective from data summary to prediction is common enough, but just check it. To take one of those simple examples that stats teachers use to illustrate a point, think of a deck of cards. There are thirteen cards in each of four suits. That is data summary. The probability that you will pick a diamond is a quarter, twenty-five percent. That is a probability assessment for the result of your decision: a prediction.

Just as a diamond is one of the cards in the deck so Ingrid is one of the population of Stockholm. But it's a bit more than that. We reasonably assume that one deck of cards is like any other. There is no reason to believe that the deck from which you choose is different from any other deck. In all cases the probability of picking a diamond is a quarter. And so we assume that the population from which Ingrid is drawn, the residents of Stockholm, is no better or worse at identifying colours than the population from which the people taking part in the experiment came. This is plausible.

~ • ~

Each row of the table is a *probability distribution*. The values in the row show how likely evidence is for each of the two alternatives, the two cars in this case. These probability distributions are called *likelihood distributions* and are the key to describing how evidence is related to alternative explanations of that evidence.

The table showed probabilities as percentages which sum to one hundred but they could also be shown as decimal fractions that sum to one, as in Table 3.2.

alternatives: car colour	*evidence*: witness statement "blue"	"green"	sum
Stig's car	0.74	0.26	1
Jan's car	0.16	0.84	1

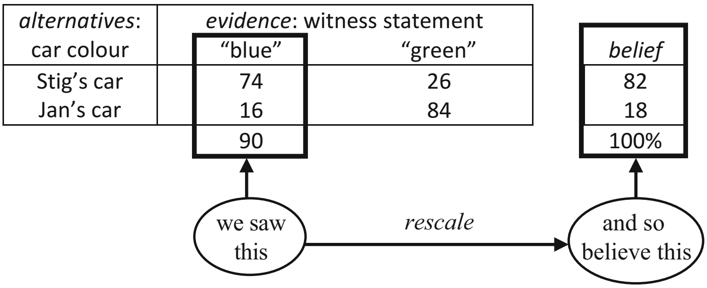
Table 3.2 Likelihood distributions

Use either. Most people find percentages easier to read though sometimes decimal fractions are easier for calculations. However you show them likelihoods are the key to using evidence.

~ • ~

Of all the possible different values of evidence, "blue" and "green", we have one, "blue". What should we believe? Do as the police did: believe in the alternatives according to how likely each is to explain the evidence. It makes sense to express this belief as a probability distribution, so rescale the likelihood values to sum to one hundred percent. Figure 3.1 shows this simple rescaling.

Fig. 3.1 What to believe given Ingrid's evidence



~ • ~

Using the word belief for these probabilities is important. It emphasises that what you believe is personal. It may be that where the evidence and likelihoods are uncontroversial we would all believe the same thing; that the probability of a coin coming down heads is fifty percent, for instance. But this is not always so. Different people may identify a different set of alternatives—Stig, Jan and Mats. Some of these differences can be accommodated in the model or extensions of it. The following chapters will show how.

~ • ~

We now have a framework for reasoning with evidence.

A simple table is all that we needed to give a structure to our problem.

The model was just "fill the rows then look at the columns". There will be a little more to it than that, but that is the essence.

Here are the components of the model as shown in the table.

1. *Each row shows an alternative.*
   What might explain the evidence; a particular deck of cards, a particular car?
   Have we missed any possible alternative explanations?

2. *Each element in a row shows a possible value of the evidence.*
   What evidence might we collect; a card, a colour?
   The list shows all possible values so that one, but only one, will be seen.

3. *Values in rows show likelihoods.*
   For each cell in a row write how likely it is that that evidence will be seen if that alternative is the true alternative explanation.
   Call these numbers, the entries in each row, the *likelihood* of each piece of evidence.
   Make sure each row sums to one hundred percent to give a complete likelihood distribution for each alternative.

4. *Rule.*
   Once the evidence is seen rescale the likelihoods in the corresponding column to sum to one hundred percent.
   The result is a probability distribution showing the degree of belief you are justified in having that each alternative is the true alternative.

The key word is **justified**.

∼ • ∼

To summarise, remember the very important rule which tells us how to make sense of what we see

believe in alternative causes or explanations in proportion to the
degree to which they explain the evidence

or **belief is proportional to likelihood**

This, in one form or another, is Bayes' Rule, named for its proposer. But just who was Bayes?

∼ • ∼

If ever you are in London go to Bunhill Fields on the edge of the City near the cluster of tech start-ups known as silicon roundabout (British irony, perhaps). Here, just off the

City Road, you find a small burial ground in which rest the remains of nonconformist churchmen. It is now a public park. Although appropriately modest this cemetery is home (if that is the right word) to William Blake and John Bunyan and Daniel Defoe. Here also is the tomb of the Bayes and Cotton families. On the top of this tomb is an inscription commemorating its 1969 restoration made possible by "subscriptions from statisticians worldwide". This generosity was in recognition of the contribution (which we have just begun to examine) of one of the Bayes family, a priest.

Fig. 3.2 Thomas Bayes (1702–1761)[1]

Thomas Bayes (Fig. 3.2) was born in 1702. His father was Joshua Bayes, a nonconformist minister at Leather Lane, in Holborn, London. In 1731, following a private education, Thomas also was ordained as a Presbyterian minister and took up a ministry at Tunbridge Wells, in Kent. He had a lifelong interest in mathematics and statistics, and was elected a Fellow of the Royal Society in 1742. He retired from the ministry in 1752 and remained in Tunbridge Wells until his death in 1761.

He published just two papers during his lifetime: in 1731, *Divine Providence and Government Is the Happiness of His Creatures*, and, five years later, *An Introduction to the Doctrine of Fluxions, and a Defense of the Analyst*, this being an attack on Bishop Berkeley following the Bishop's attack on Newton's calculus.

After Bayes' death his friend Richard Price, also a Bunhill Fields resident, sent a further paper by Bayes to the Royal Society and in 1763 it was published in the *Philosophical Transactions of the Royal Society of London* as *Essay Towards Solving a Problem in the Doctrine of Chances*. It is this paper which contains the famous rule.[2]

∼ • ∼

[1]Original source unknown. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Thomas_Bayes.gif. Accessed 16 September 2017.

[2]There are a number of accounts of Bayes life and work. Here are two: Barnard [1] and Dale [2]. Another, which concentrates more on the post-war spread of Bayes' ideas, is Fienberg [3].