Gerald's Column
by Gerald Fitton

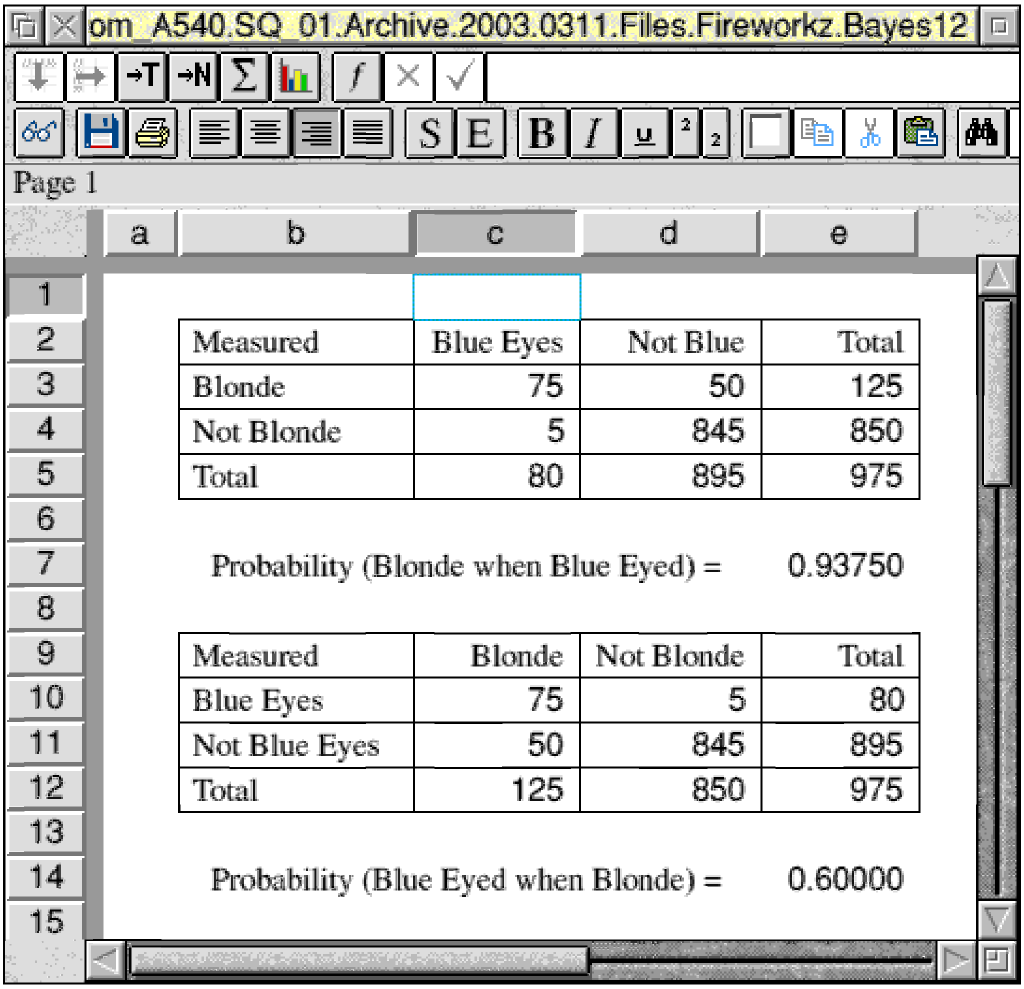
This is my third article about Bayes' Theorem. In my two previous articles I introduced two slightly different examples and, in this month's article, I had intended to introduce a third variant. However, my first article has generated an overwhelming amount of correspondence some of which requires a public rather than individual response. I blame Paul! It was he who added the postscript to my article on page 27 in Vol 16, No 12!

The worked example I shall use for this third variant will be its application to the detection of spam emails. It will have to wait until I have answered those questions arising from my first two articles which I believe will be of benefit to all Bayes' Theorem aficionados.

Hypothesis or Event

My first set of questions can be summarised as:
"What is the difference between a Hypothesis and an Event?"

I believe that it was the nature of the example I used to introduce Bayes' Theorem which has prompted this question. My only defence is that I wanted to introduce the logic behind the formula in a gentle way rather than jump in with both feet and then, as is so often the case with explanations of Bayes' Theorem, drown you rather than teach you to swim!



Measured	Blue Eyes	Not Blue	Total
Blonde	75	50	125
Not Blonde	5	845	850
Total	80	895	975

Probability (Blonde when Blue Eyed) = 0.93750

Measured	Blonde	Not Blonde	Total
Blue Eyes	75	5	80
Not Blue Eyes	50	845	895
Total	125	850	975

Probability (Blue Eyed when Blonde) = 0.60000

The mistake which I made was to provide you with a complete set of raw data rather than emphasising that it might be only a sample of data from which we had to draw general conclusions. Nevertheless, let's continue to use this example.

Look at the screen shot above. It is derived from my 'Blue Eyed Blonde' example. Now ask yourself the following question: "Which of the two attributes, 'Blue Eyes' or 'Blonde' is the Event and which is the Hypothesis?"

The answer to this question is: "It doesn't matter. You can chose either of them!"

From the Known to the Unknown

In general terms the Hypothesis is the thing you don't know or are unsure about. The Event is something which has happened so you do know the result. In my example we know everything. Using the statistical jargon, we have access to the 'Raw Data'.

As many of you (including our greatly respected puzzle specialist, Colin Singleton) pointed out to me, because we have direct access to the raw data there is absolutely no need to use the sledge hammer of Bayes' Theorem to crack this particular walnut.

If we choose as our unknown attribute "Blonde" and as our known attribute "Blue Eyed" then, in terms of Bayes' Theorem, it is "Blonde" that is the Hypothesis and "Blue Eyed" is the Event. The relevant question is: "Knowing that this person is Blue Eyed what is the probability they're also Blonde?"   The answer is in e7 = d3/d5 = 94% (approximately).

Now turn your attention to the lower half of the screenshot. All that I have done is to transpose the first table to create the second. Fireworkz contains transpose not only as a command but also as a function. I entered the formula set_value(a8e12, transpose(a1e5)) into cell e8. Then I deleted the formula from e8. Easy, isn't it?

You can do exactly the same thing using PipeDream.

The value in e14 = d10/d12 = 60%.

In this second half of the screenshot we can regard the attribute "Blonde" as the known Event and the unknown attribute "Blue Eyes" as the Hypothesis.

Question: Which is the Hypothesis and which the Event? Answer: You can chose either! You can choose either because you have access to the raw data.

Rather than simply regarding the second table as a valid mathematical transposition of the first table table let us consider two different scenarios.

In the first scenario I want to select someone who is Blonde for a part in my stage play. On the questionnaire the (incompetent) agency have listed the eye colour for all the candidates but omitted the highly important hair colour! Time is limited. We can't interview everyone, how can we shorten the list? The answer is that we will increase our chances of finding a Blonde from about 13% to 94% by restricting ourselves to those candidates who have the attribute "Blue Eyes". Instead of one chance in eight we are now fairly certain (about 19 out of 20) that we'll find a Blonde as our first choice! That's a big improvement.

I shall leave you to construct the second scenario. It is a different scenario, isn't it?

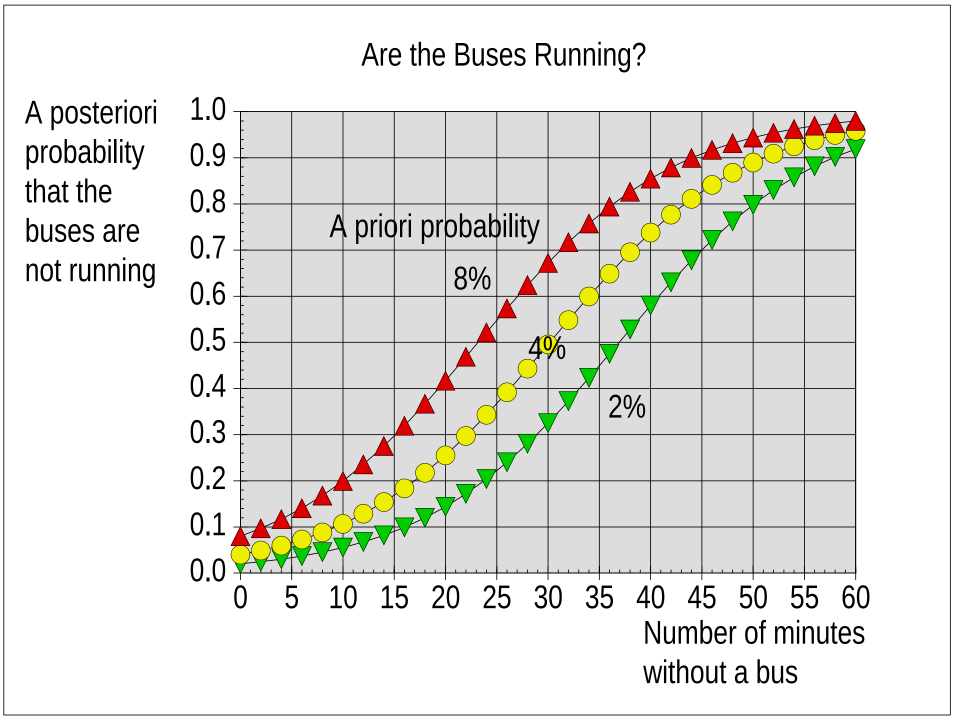
The Unknowable Probability

In my second article the distinction between Hypothesis and Event is much clearer. As you will recall, my objective was to catch a bus into town and thereby avoid the fine payable on my library book. The bus never came!

There are three Hypotheses. They are: H_2, the buses run on time, H_3, that the bus arrival times are random and H_1, that neither of these mathematical models are sufficiently valid for my purpose. Indeed, only when my friend with the car told me that the buses were not running did it become certain that H_1 alone was valid.

The feature which distinguishes these three Hypotheses from the sequence of missing bus Events is that we don't know which Hypothesis is most applicable to the situation. It could be any of the three. It can not be more than one of them. They are 'Mutually Exclusive'.

It is important that you recognise that for this example there is no raw data to fall back on. Even more important is that you realise that there is absolutely no way of generating a full set of raw data on the day I want to catch a bus.



Above all, not only are the three 'A Priori' probabilities, $P(H_1)$, $P(H_2)$ and $P(H_3)$, unknown, what is of paramount importance for you to realise is that they are unknowable. We have to resort to either guesswork or prejudice when deciding what values for these 'A Priori' probabilities we shall insert into Bayes' Formula.

Although Bayes' Theorem has little if anything to say about these unknowable 'A Priori' probabilities which are an essential ingredient of Bayes' Formula, we should not make the assumption that the Reverend Thomas Bayes was unaware of this problem. Not only was he a 16th Century Presbyterian Minister and Mathematician but he was also a Philosopher.

It is to Bayesian Epistemology, a branch a Philosophy, which we must turn for further enlightenment. The Reverend Thomas' paper entitled "Essay towards solving a problem in the doctrine of chances" contains the essence of Bayesian Epistemology.

What does 'Probability' mean?

I have not been asked this question directly. Rather than provide individual answers to the myriad of subtly different, but related, questions which have arrived in my email inbox, I shall choose to answer this one (at least in part) for I believe it is the question which should have been asked—and perhaps I should have addressed sooner.

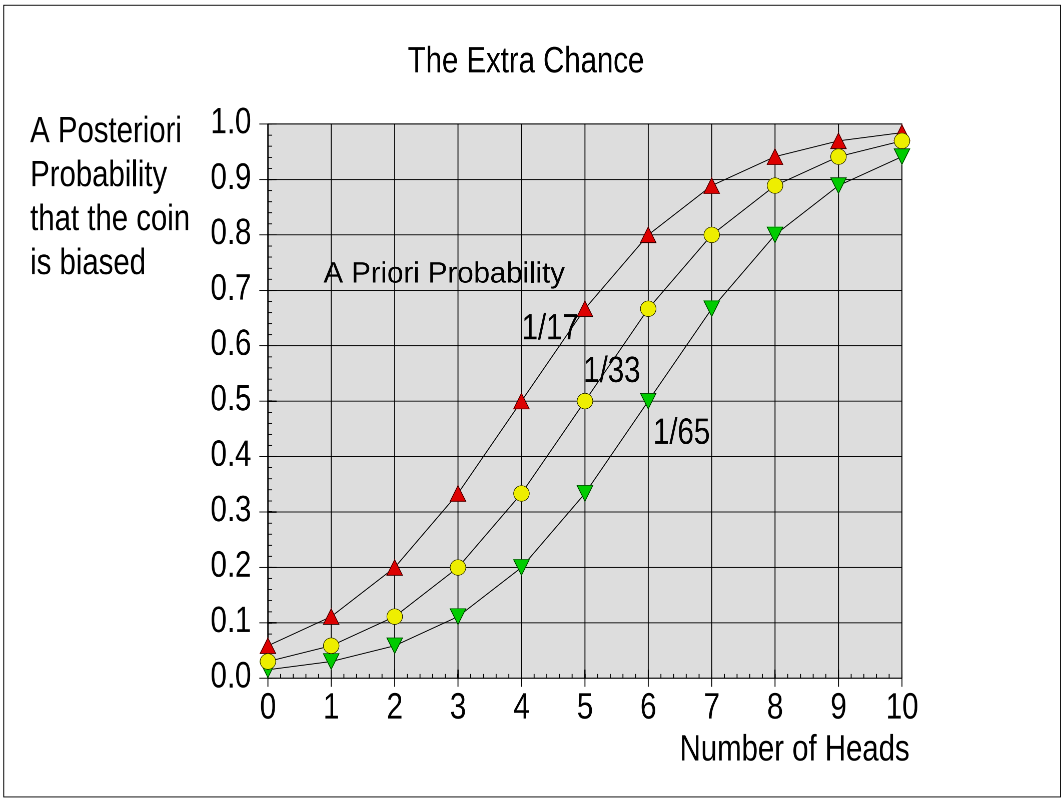
The Unbiased Coin

You will recall my flippin' coin. That coin was not an unbiased coin.
The coin I want you to consider now is a truly hypothetical, totally unbiased coin.

I ask you to consider this question: "What is the probability that it will come up 'Tails'?"
The answer, which will be agreed by all rational people, is an absolutely solid 50%!
In this totally hypothetical context the word "Probability" has a definable meaning because we (being rational) can all agree on its value. We might say that this probability has a value which is objective rather than subjective. That value is undoubtedly 50%.

The Flippin' Coin

The flippin' coin experiment I ran with you is not hypothetical. You experienced its reality.



Look at the chart. You will have to wait until next month if you wish to know why this PipeDream chart has the cryptic name “The Extra Chance” but here is a clue. The ‘A Priori’ probabilities for the three lines are the fractions 1/17, 1/33 and 1/65. The number of Heads required at which the ‘A Posteriori’ Probability hits (exactly) 50% are 4, 5 and 6 respectively. Write to me if you can see the (obscure?) connection.

I produced the chart using PipeDream. I use PipeDream in preference to Fireworkz for charts because it is much more flexible (once you get to know all the facilities hidden in the sub menus). The chart is totally ‘live’. By this I mean that when I vary the data the chart will be redrawn automatically. I can even change the text, the position of the text and even the size of the text with great ease.

One of the questions which I receive regularly is how a graph like this with, say, three lines can be produced using PipeDream. In Fireworkz it is a bit easier but then, the Fireworkz charting package does not have as many features as PipeDream does. This month I have space only for a clue and not an explanation. My clue is <Ctrl CHA> (CHart Add).

Conditional Probability

Enough of this spreadsheet digression. Back to the chase.
In the context of my flippin’ coin what does, what can the word “Probability” mean?
For example: “What is the correct ‘A Priori’ Probability that the coin is unbiased?”

I’ve seen many text books which tackle this question by introducing an interesting but difficult concept called “Conditional Probability”. The difficulty which students have with this concept is that they try hard to convince themselves that this “Conditional Probability” has an absolute objective value rather like the probability of the totally and utterly hypothetical, non existent, unbiased coin. Like the man (in my day traditionally Irish) giving directions to the lost motorist would have said, “I wouldn’t have started from here!”

Which Probabilities are we sure of?

Here is a reminder of the two Hypothesis version of Bayes’ Formula.
$P(H_1|E_1) = P(H_1) * P(E_1|H_1) / (P(H_1) * P(E_1|H_1) + P(H_2) * P(E_1|H_2))$.

It is common practice to call all these P things probabilities. $P(H_1)$ and $P(H_2)$ are common or garden probabilities whereas $P(E_1|H_1)$, $P(E_1|H_2)$ and $P(H_1|E_1)$ are of the new fangled “Conditional Probability” type!

This distinction is illusory!
Making this distinction leads to confusion.
Let’s not go down that road.
Taking the Irishman’s wise advice to the lost motorist, let’s start somewhere else.

Some of these probabilities can be calculated by making assumptions (based on symmetry—remember?) about the Probability Distribution Function (PDF) or they can be estimated from measurements (experimental data). In previous articles I have discussed the construction of PDFs using symmetry and by analysing experimental data.

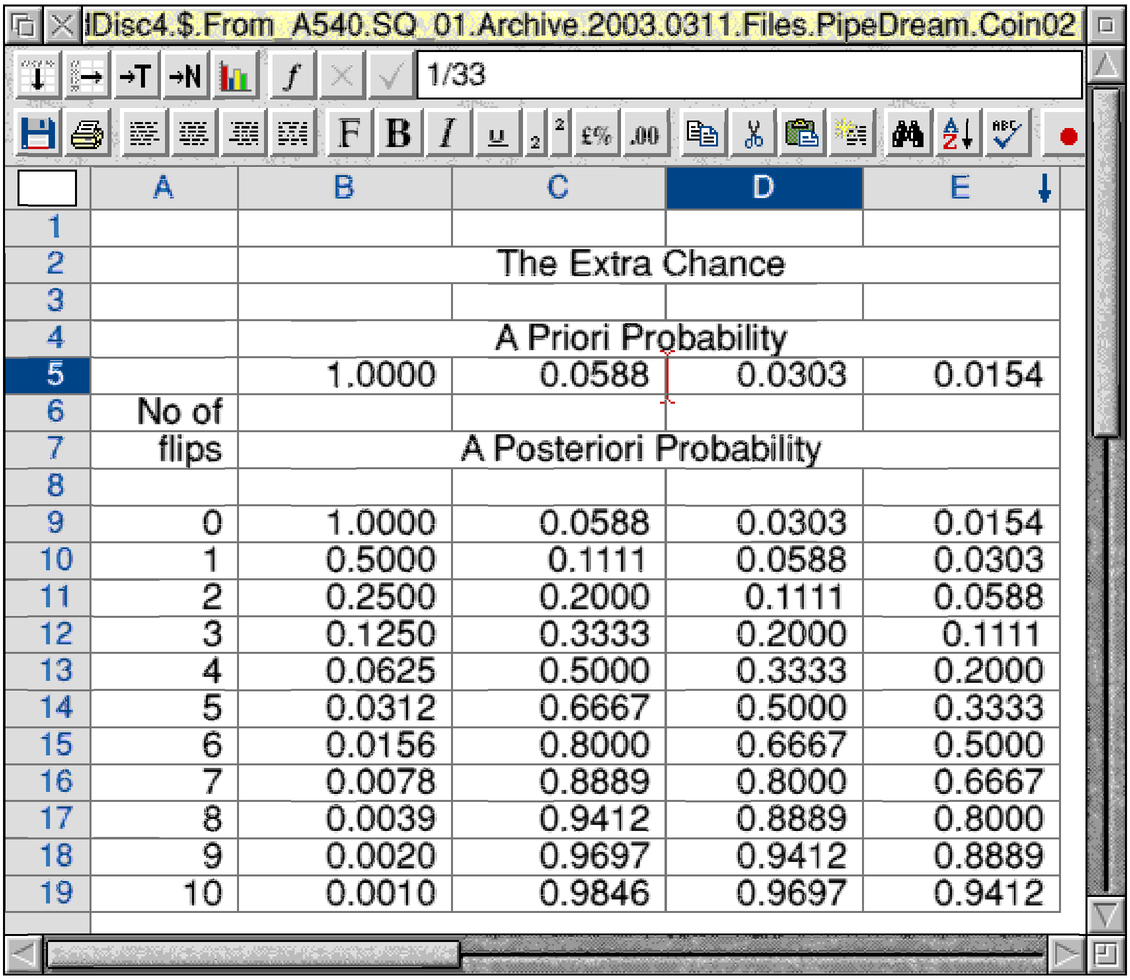
The two probabilities which can be calculated or estimated are $P(E_1|H_1)$ and $P(E_1|H_2)$.

In the case of the flippin' coin:
Hypothesis 2, H_2, is that the coin is unbiased.
$P(E_1|H_2)$ is the probability that the coin will come down tails if the coin is unbiased.
There is no doubt that we will all agree that this has the objective value of 50%.
Hypothesis 1, H_1, is that the coin is biased.
I shall avoid discussing the extent of the bias and stick with my flippin' coin.
You'll recall that it is a double headed half crown.
$P(E_1|H_1)$ is the probability that the coin will come down tails if the coin is biased.
There is no doubt that we will all agree that this has the objective value of 0%.

That's the easy bit. Now the hard bit.

The only values we don't know on the right hand side of the Bayes' Theorem Formula are $P(H_1)$ and $P(H_2)$. Furthermore we know that $P(H_1) + P(H_2) = 1$ because between H_1 and H_2 we must exhaust all possibilities. Once we know $P(H_1)$ we can calculate $P(H_2)$.

The numerical value of $P(H_1)$ might be written in words as "The probability that the coin is biased". How can we find a value for $P(H_1)$, this missing probability? Of course, with hindsight we know now that the coin is a double headed half crown but you didn't know that until I told you so. That information is unavailable to you. So let me repeat my question: "What value shall we use for $P(H_1)$, the probability that the coin is biased?"



	A	B	C	D	E
1					
2		The Extra Chance			
3					
4		A Priori Probability			
5		1.0000	0.0588	0.0303	0.0154
6	No of				
7	flips	A Posteriori Probability			
8					
9	0	1.0000	0.0588	0.0303	0.0154
10	1	0.5000	0.1111	0.0588	0.0303
11	2	0.2500	0.2000	0.1111	0.0588
12	3	0.1250	0.3333	0.2000	0.1111
13	4	0.0625	0.5000	0.3333	0.2000
14	5	0.0312	0.6667	0.5000	0.3333
15	6	0.0156	0.8000	0.6667	0.5000
16	7	0.0078	0.8889	0.8000	0.6667
17	8	0.0039	0.9412	0.8889	0.8000
18	9	0.0020	0.9697	0.9412	0.8889
19	10	0.0010	0.9846	0.9697	0.9412

This is the point at which Bayesian Epistemology comes to our rescue.
This is the point at which those who like to believe that they are 'rational' often get lost!

'A Priori' Probability

Let me repeat that it is to Bayesian Epistemology, a branch a Philosophy, to which we must turn if we wish to enter a value for $P(H_1)$ in the Bayes' Theorem Formula. It is in the finer detail of the Reverend Thomas' paper entitled "Essay towards solving a problem in the doctrine of chances" that we we shall find a rational methodology for choosing $P(H_1)$.

What is the great wisdom that the devout Presbyterian Reverend reveals in his "Essay"?
What is the great revelation which has lead to the birth of Bayesian Epistemology?

This rational, knowledgeable, devout and inspired thinker provided the answer.
It was a one word answer. What was the word?

It was "Guess"!

By "Guess" I am not asking you to guess what he wrote. No! I am telling you what he wrote (even though I am not using the words he used). The most logical and rational thing you can do about a value for $P(H_1)$ is make a guess as to what it might be!

Now look! There is no way that this sort of Probability can be considered to be an objective number such as the 50% which we, being rational people, can all agree is the correct value for $P(E_1|H_2)$. If it is a probability at all then the meaning is not the same as the word "Probability" when applied to the unbiased coin. Indeed, the use of the word "Probability" is inappropriate for the value of $P(H_1)$.

I think only Mathematicians using the Bayes Theorem Formula still call it a Probability.
The followers of Bayesian Epistemology prefer to use the phrase "Degree of Belief".
In more auspicious (and older) texts the Latin expression 'A Priori' is used.

Conclusion

Once we leave the safe world of hypothetical mathematical models (such as the impractical unbiased coin) and Samples which we assume are representative of a Population we are studying, then we are reduced to guesswork.

Although we shall never know whether a Hypothesis is true or false there is no doubt that Bayes' Formula is a useful practical tool. It is this formula which allows us to convert guessed at 'A Priori' probabilities into more usable 'A Posteriori' probabilities.

As we learn, these 'A Posteriori' probabilities, $P(H_1|E_1)$, (usually) tend towards either zero or one (impossibility or certainty). Values of probability close to zero or one are much more usable than the mediocre and arbitrary 'A Priori' probabilities. In philosophical jargon, Bayes' Formula forms the rational basis for a change in our 'Degree of Belief' about uncertain Hypotheses. It is our experience (Events) which drives this change in our 'Degree of Belief'. Scientific advancement is founded on Bayesian Epistemology.

Communication

Please contact me by email (preferred) or by letter if you have any questions or comments.

Because of spam problems I shall shortly set up a Spam Filter which will reject (bounce) emails directed to <gerald@abacusline.demon.co.uk> unless you are on my personal ‘white list’. Please email me at <Archive@abacusline.demon.co.uk> if your email relates to articles in the Archive magazine or at <GoldLine@abacusline.demon.co.uk> if your correspondence relates to PipeDream, Fireworkz or GoldLine or, if your correspondence relates to Living with Technology, then please use <LwT@abacusline.demon.co.uk>.

Note to Paul: Please change the email address in the Fact File to Archive@abacusline...