Massimo Fuggetta
www.massimofuggetta.com

October 2013
Last revision: March 2014

BLINDED BY EVIDENCE

'Is the Pope Italian?' is a common expression used to remark on the obvious. In fact, since the beginning, almost 80% of Popes have been Italian[1]. Not entirely obvious, then (especially in recent decades), but highly likely. Take a Pope: the probability that he is Italian is 80%.

Now take an Italian: what is the probability that he is a Pope? Unless you ask his mother, it is much lower. In statistical parlance, the probability that someone is Italian, given that he is a Pope, is definitely not the same as the probability that he is a Pope, given that he is Italian.

As obvious as this appears, people regularly confuse the probability of a hypothesis, given some evidence, with the probability of the evidence, given the hypothesis. It is a well-known phenomenon, which psychologists call, among other things, the *Inverse Fallacy*.

This paper contains an extensive analysis of the Inverse Fallacy. Its main claim is that the fallacy is best seen as a Prior Indifference Fallacy: the unwarranted and generally erroneous assumption that the prior probability of a hypothesis is 50%, i.e. the hypothesis is equally likely to be true or false. Seeing the Inverse Fallacy as prior indifference sheds light on what it is and what it isn't, why it arises and persists, and how it can be avoided.

Section 1 illustrates the Inverse Fallacy through a stylized example. Section 2 introduces the Bayes' Theorem and defines the main concepts used throughout the paper. Section 3 defines types of evidence and describes the iterative nature of Bayesian updating. Section 4 introduces and discusses the Prior Indifference Fallacy. Sections 5 and 6 examine the fallacy in the different contexts of hard and soft evidence. Section 7 relates the fallacy to Knightian uncertainty and ambiguity aversion. Section 8 shows how prior indifference underlies three main cognitive heuristics: representativeness, anchoring and availability. The final section concludes the paper.

1. A disturbing test

You hear on television that forty people have recently died from a lethal virus[2]. You actually knew one of the deceased. Although you are not the impressionable type, just to be sure you call your doctor. The doctor says you shouldn't worry but, to be safe, you can take a test that is 99% accurate at spotting the virus: if you have it, it will tell you so with near certainty. Well, let's do it then, you say, and fix an appointment for the next day.

[1] Wikipedia, Popes by Nationality.

[2] This a dramatized version of the Harvard Medical School test presented in Casscells et al. (1978) and discussed in Tversky, Kahneman (1982).

That night you can't sleep. Your mind keeps going back to that poor fellow who died, just 32, leaving a wife and two children. You have no reason to worry, but what if you have the virus? And, much worse, what if you have it, but you happen to be in that 1% of cases for which the test is wrong? You think you don't have the virus but you actually have it: that would be just horrible. So the next day you ask your doctor whether there is an even better test that could put your mind completely at rest. Yes – says the doctor – there is one, but it is very expensive and very painful. However, if you have the virus, this test will tell you with 100% certainty. Wait a minute, doctor – you say – is this really a perfect test? You are saying that, if I have the virus, the test will spot it with 100% accuracy. But what if I *don't* have the virus? How accurate is the test in that case? If you don't have the virus – says the doctor – the test will correctly tell you so with 95% accuracy. So it is not perfect, but it is still very accurate.

You take a big breath and decide to go for it. Back home, you spend another sleepless night – half from the lingering pain, half from the emotional turmoil. The next morning you rush to the doctor's clinic to get the result and put an end to this mental and physical torture. Immediately you notice there is something wrong on the doctor's face. In a highly embarrassed tone, the doctor gives you the verdict: you tested positive – the test said you have the virus. Remember, there is still a chance that the test is wrong, but I am afraid it is not very high. Sorry.

Desperate, you stagger back home and start writing your will, when your friend Thomas, the statistician, calls you on the phone. You tell him the awful news but, to your dismay, he starts laughing. So what? – he says – I know this test. Do you want to know the probability that you have the virus? Yes, about 100% – you cry. Think again – says Thomas – the probability you have the virus is less than 2%. I'll come to your place and explain – if you offer me a beer or two.

What happened? If you are like most people, you are very confused – and very interested to hear the statistician's explanation. So here it is. The doctor said that, if you have the virus, you would test positive with 100% certainty. We can write this as P(+|V)=1, which reads: the probability of testing positive, given that you have the virus, is 100%. In answer to your question, he also told you that, if you don't have the virus, the test is still very accurate, although not infallible. We can write this as P(–|no V)=0.95, which reads: the probability of testing negative, given that you don't have the virus, is 95%. What the doctor didn't tell you – says Thomas – is that the virus is rare: it hits only one out of a thousand people. So what? – you say – however rare it might be, the test is saying that I have it, and the test is very accurate. Not so – says Thomas – you need to know how rare the virus is in order to work out what you are really after: the probability of having the virus, given that you tested positive: P(V|+). To calculate this probability, Thomas writes down a formula uncovered by his 18th century namesake, Reverend Thomas Bayes, which says:

$$P(V|+)=\frac{P(+|V)P(V)}{P(+)} \tag{1}$$

The doctor told you P(+|V)=1 and P(–|no V)=0.95, and you mistakenly thought this meant that P(V|+) was very high. This is the *Inverse Fallacy*: you confused the probability of the hypothesis, given some evidence, with the probability of the evidence, given the hypothesis. But you can easily correct your mistake: Thomas told you that the virus has a probability of 1/1000, so P(V)=0.001. He now tells you how to calculate P(+), the probability that you test positive: P(+)=P(+|V)P(V)+P(+|no V)P(no V)=1×0.001+0.05×0.999=0.051. That is all you need to calculate P(V|+) and, to your great relief, the answer is 0.0196, i.e. less than 2%. Your expectation of certain death just turned into a 98% chance of survival.

2. The general case

To see what is happening, let's analyse the general case of a hypothesis H, which can be either true or false. The probability that H is true is P(H) and the probability that it is false is 1-P(H).

Empirical knowledge consists in the accumulation of evidence in order to evaluate hypotheses. Any sign that can be related to a hypothesis is a form of evidence about the hypothesis. When the sign is present, we say that evidence is positive, with probability P(E). When the sign is absent, we say that evidence is negative, with probability 1-P(E). We continuously revise the probability of hypotheses in the light of new evidence. There are four possible cases:

	H is true	H is false
E is positive	True Positives	False Positives
E is negative	False Negatives	True Negatives

Through direct observation or by other means, we form beliefs about the probabilities of the four cases:

True Positive Rate: probability of positive evidence, given that the hypothesis is true: P(E|H).

False Positive Rate: probability of positive evidence, given that the hypothesis is false: P(E|not H).

True Negative Rate: probability of negative evidence, given that the hypothesis is false: P(not E|not H).

False Negative Rate: probability of negative evidence, given that the hypothesis is true: P(not E|H).

These conditional probabilities can be represented as in the following table:

Table 1 **Anterior probabilities**

	H is true	H is false	TOTAL
E is positive	P(E|H)	P(E|not H)	?
E is negative	P(not E|H)	P(not E|not H)	?
TOTAL	100%	100%	

The probabilities in Table 1 measure the ex-ante *accuracy* of the evidence and are therefore called anterior probabilities.

For example, if the hypothesis is: There is a fire, the evidence may be: There is smoke. Evidence can give two right responses: True Positives (smoke, fire) and True Negatives (no smoke, no fire) and two wrong responses: False Positives (smoke, no fire) and False Negatives (no smoke, fire). False Negatives, i.e. wrongful rejections of the hypothesis, are known as Type I errors, while False Positives, i.e. wrongful acceptances of the hypothesis, are known as Type II errors. Ideally, we would like both errors to have the smallest probabilities, but typically there is a trade-off between the two. At the extremes, never rejecting the hypothesis would entirely avoid Type I errors, but it would likely lead to a larger probability of Type II errors. Vice versa, always rejecting the hypothesis would eliminate Type II errors, but entail a higher probability of Type I errors.

Notice that, while the columns in Table 1 must add up to 1 (since evidence is either positive or negative) the rows don't have to: there is no reason for the two errors to have equal probabilities. We call evidence in which the two probabilities happen to be equal – hence the rows of Table 1 also add up to 1 – *symmetric evidence*. This may well be a natural occurrence, but is not generally true. For instance, in our virus test the probability of Type I errors is zero, while the probability of Type II errors is 5% – i.e. evidence is not symmetric.

Anterior probabilities define the accuracy of the evidence in favour or against H. But, as our virus story shows, the probabilities we are ultimately interested in are the ones in Table 2:

Table 2 **Posterior probabilities**

	H is true	H is false	TOTAL
E is positive	P(H|E)	P(not H|E)	100%
E is negative	P(H|not E)	P(not H|not E)	100%
TOTAL	?	?	

These are called posterior probabilities, as they measure the probability of the hypothesis after the arrival of new evidence. They define the *support* of the evidence in favour or against the hypothesis. We know from (1) how to calculate P(H|E). The other three probabilities can be calculated using the same method. Notice that in Table 2 it is the rows that must add up to 1, whereas the columns may not sum to 1, again unless the evidence is symmetric.

In the sequel of the paper, we shall adopt the following notation:

Table 3 **Notation**

P(H)	Base Rate, Unconditional, Prior Probability	BR
P(E|H)	True Positive Rate, Likelihood, Sensitivity, Hit Rate	TPR
P(not E|H)	False Negative Rate, Probability of Type I error, Miss Rate	FNR=1-TPR
P(E|not H)	False Positive Rate, Probability of Type II error, False Alarm Rate	FPR
P(not E|not H)	True Negative Rate, Power, Specificity	TNR=1-FPR
P(H|E)	Probability that H is true, given positive evidence	PP
P(not H|E)	Probability that H is false, given positive evidence	1-PP
P(H|not E)	Probability that H is true, given negative evidence	NP
P(not H|not E)	Probability that H is false, given negative evidence	1-NP

The first column in Table 3 gives the common mathematical notation of the probabilities; the middle column gives definitions and other terms used to denote them; the third column indicates the notation that, for ease of exposition, we shall henceforth use in the paper.

Using our notation, we can write Bayes' Theorem as:

$$\mathrm{PP}=\frac{\mathrm{TPR}\cdot\mathrm{BR}}{\mathrm{TPR}\cdot\mathrm{BR}+\mathrm{FPR}\cdot(1-\mathrm{BR})}$$

$$\mathrm{NP}=\frac{(1-\mathrm{TPR})\cdot\mathrm{BR}}{(1-\mathrm{TPR})\cdot\mathrm{BR}+(1-\mathrm{FPR})\cdot(1-\mathrm{BR})}=\frac{\mathrm{FNR}\cdot\mathrm{BR}}{\mathrm{FNR}\cdot\mathrm{BR}+\mathrm{TNR}\cdot(1-\mathrm{BR})} \tag{2}$$

PP is the posterior probability we are interested in. In the virus story, it is the probability that you have the virus, given that you tested positive. PP depends on BR, TPR and FPR. BR, the unconditional probability of H, is known as the *Base Rate*, or *Prior Probability* of the hypothesis. TPR is the *True Positive Rate*, also known as *Likelihood*, *Sensitivity*, or *Hit Rate*. The probability of Type I errors is FNR, known as the *False Negative Rate*, or *Miss Rate*. FPR is the probability of Type II errors, known as the *False Positive Rate*, or *False Alarm Rate*. TNR is the *True Negative Rate*, also known as *Power* or *Specificity*. Likewise, NP is the probability that you have the virus, given that you tested negative. NP also depends on BR, TPR and FPR.

Accuracy is equal to the average of the True Positive Rate and the True Negative Rate:

$$\mathrm{A}=(\mathrm{TPR}+\mathrm{TNR})/2=0.5+(\mathrm{TPR}\text{-}\mathrm{FPR})/2 \tag{3}$$

Perfect accuracy has TPR=1 and FPR=0, hence A=1. Coin-toss accuracy, i.e. perfect inaccuracy, has TPR=FPR, hence A=0.5. Perfect contrary accuracy has TPR=0 and FPR=1, hence A=0. Notice that, if evidence is symmetric (FPR=FNR), then A=TPR.

(2) can be rewritten as:

$$\frac{\mathrm{PP}}{1-\mathrm{PP}}=\frac{\mathrm{TPR}}{\mathrm{FPR}}\cdot\frac{\mathrm{BR}}{1-\mathrm{BR}}$$

$$\frac{\mathrm{NP}}{1-\mathrm{NP}}=\frac{\mathrm{FNR}}{\mathrm{TNR}}\cdot\frac{\mathrm{BR}}{1-\mathrm{BR}} \tag{4}$$

(4) is known as Bayes' Theorem in *odds form*. Odds are the ratio between the probability that H is true and the probability that it is false. PP/(1-PP) are the Posterior Odds of H, given positive evidence. BR/(1-BR) are the prior (or Base) Odds of H. TPR/FPR is the *Likelihood Ratio*. Hence we can write PO=LR·BO: Posterior Odds are a linear function of Prior Odds, with slope LR. The Likelihood Ratio transforms Prior Odds into Posterior Odds. Likewise, NP/(1-NP) are the Posterior Odds of H, given negative evidence, with Likelihood Ratio FNR/TNR.

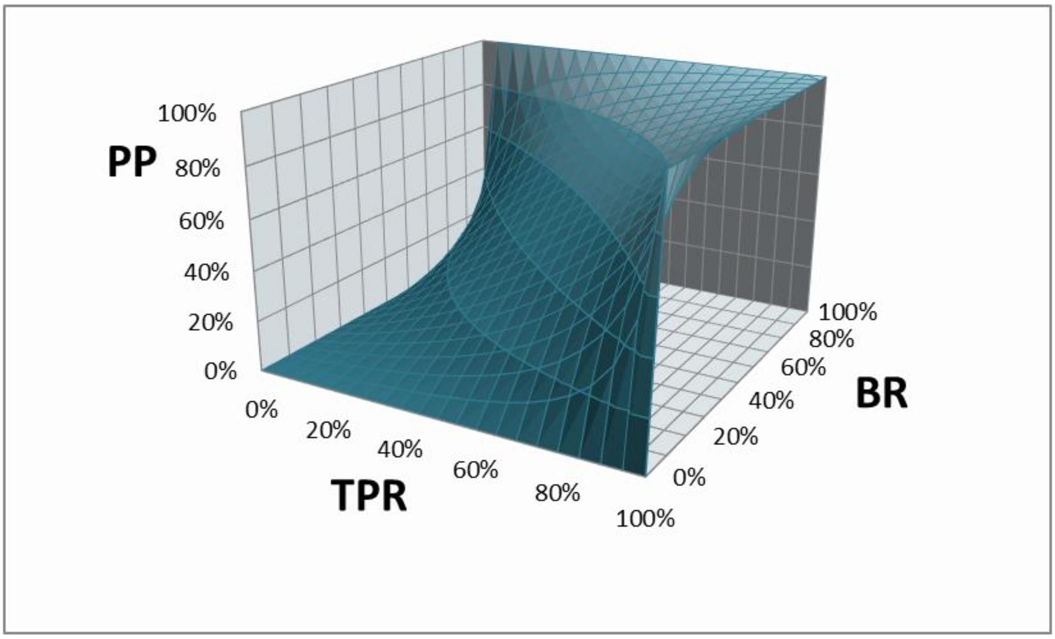
In case of symmetric evidence, (2) becomes:

$$\mathrm{PP}=\frac{\mathrm{TPR}\cdot\mathrm{BR}}{\mathrm{TPR}\cdot(2\cdot\mathrm{BR}-1)+1-\mathrm{BR}}$$

$$\mathrm{NP}=\frac{\mathrm{FNR}\cdot\mathrm{BR}}{\mathrm{FNR}\cdot(2\cdot\mathrm{BR}-1)+1-\mathrm{BR}} \tag{5}$$

PP in (5) can be seen graphically in Figure 1.

Figure 1 – Relationship between posterior, anterior and prior probabilities for symmetric evidence



3. Types of evidence

The probability pair (TPR,FPR) defines accuracy which, together with the Base Rate BR, determines the probability of the hypothesis in the light of the evidence.

From (2), if BR=1 then PP=NP=1, irrespective of TPR and FPR. We call this *Faith*: a prior belief in the truth of the hypothesis, which requires no evidence and which no amount of evidence, however strong, can change. Likewise, if BR=0 then PP=NP=0, irrespective of TPR and FPR. This is Faith in the falsity of the hypothesis, again irrespective of any evidence.

But PP and NP can reach the boundaries of the probability spectrum also *as a result* of evidence. We call this *Certainty*. Certainty can result from perfect or conclusive evidence.

Perfect Evidence is defined as TPR=1 and FPR=0. It is evidence incompatible with False Negatives and False Positives. From (2), PP=1 and NP=0, irrespective of BR. With positive evidence, the hypothesis must be true; with negative evidence, it must be false. From (3), perfect evidence is perfectly accurate: A=1. Likewise, perfect *contrary* evidence is defined as TPR=0 and FPR=1. It is evidence incompatible with True Positives and True Negatives. From (2), PP=0 and NP=1, again irrespective of BR. With negative evidence, the hypothesis must be true; with positive evidence, it must be false. Perfect contrary evidence is perfectly contrarily accurate: A=0.

Perfect evidence is conclusive: it transforms subjective beliefs into objective, prior-free Certainty. But evidence does not need to be perfect in order to be conclusive. Imperfect, c*onclusive evidence* is defined as TPR=1 *or* FPR=0 (but not both). Alternatively, as TPR=0 *or* FPR=1 (but not both). There are four possible cases[3]:

[3] The four cases get their name from the hypothesis of guilt. A Smoking Gun is conclusive positive evidence that the suspect is guilty. A Perfect Alibi is conclusive positive evidence that the suspect is innocent.

a) *Smoking Gun*: FPR=0. It is positive evidence incompatible with False Positives. From (2), PP=1, irrespective of BR and TPR. With a Smoking Gun, the hypothesis must be true. However, since FNR>0, then NP>0: without a Smoking Gun, the hypothesis is not necessarily false.

b) *Perfect Alibi*: TPR=0. It is positive evidence incompatible with True Positives. From (2), PP=0, irrespective of BR and FPR. With a Perfect Alibi, the hypothesis must be false. However, since TNR>0, then NP<1: without a Perfect Alibi, the hypothesis is not necessarily true.

c) *Barking Dog*: FNR=0. It is negative evidence incompatible with False Negatives. From (2), NP=0, irrespective of BR and TNR. Without a Barking Dog, the hypothesis must be false. However, since FPR>0, then PP<1: with a Barking Dog, the hypothesis is not necessarily true.

d) *Strangler Tie*: TNR=0. It is negative evidence incompatible with True Negatives. From (2), NP=1, irrespective if BR and FNR. Without a Strangler Tie, the hypothesis must be true. However, since TPR>0, then PP>0: with a Strangler Tie, the hypothesis is not necessarily false.

The four cases are summarized in Table 4:

Table 4 **Conclusive Evidence**

	H is true	H is false
E is positive	Smoking Gun: FPR=0, FNR>0 PP=1, NP>0	Perfect Alibi: TPR=0, TNR>0 PP=0, NP<1
E is negative	Strangler Tie: TNR=0, TPR>0 NP=1, PP>0	Barking Dog: FNR=0, FPR>0 NP=0, PP<1

Faith and Certainty drive probability to one of the two boundaries of its spectrum. But whereas Faith requires no evidence, Certainty is entirely based on conclusive evidence. And whereas Faith relies on extreme priors, conclusive evidence renders priors irrelevant. This is the allure of conclusive evidence: it frees our beliefs from subjective priors. Whatever we thought beforehand, the acquisition of conclusive evidence implies that

Conclusive evidence is often used in works of fiction to bring out final certainty. Sherlock Holmes is the supremo of conclusive evidence. His incessant accumulation of evidence often culminates with a conclusive piece, thanks to which his deductions about guilt or innocence leave the realm of probability and, through inescapable logic, reach the pinnacle of certainty. In *Silver Blaze*, Sherlock Holmes proves that Simpson could not have killed Straker, because the dog didn't bark:

> "Is there any point to which you would wish to draw my attention?"
> "To the curious incident of the dog in the night-time."
> "The dog did nothing in the night-time."
> "That was the curious incident," remarked Sherlock Holmes.

A Barking Dog is conclusive negative evidence that the suspect is innocent.

In many of Alfred Hitchcock's movies, the main character is an innocent man, being cornered by an accumulation of circumstantial evidence pointing to his guilt, until a single piece of conclusive evidence proves his innocence. In the final scene of Hitchcock's *Frenzy*, Inspector Oxford nails down the Covent Garden strangler: "Mr. Rusk, you're not wearing your tie", thus proving that Dick Blaney – until then the chief suspect – is innocent.

A Strangler Tie is conclusive negative evidence that the suspect is guilty.

the hypothesis *must* be true (or false). However, conclusive evidence should not be confused with perfect evidence. A Smoking Gun proves that the suspect must be guilty, irrespective of our priors. But it is wrong to conclude that, if no Smoking Gun is found, the suspect must be innocent. Whether we believe he is innocent or not continues to depend on our priors.

If, on the other hand, TPR and FPR are not 0 or 1, but lie somewhere between the two boundaries, evidence is *inconclusive*. With inconclusive evidence, beliefs cannot disengage from priors and can only approximate but never reach Certainty.

Inconclusive evidence is *confirmative* if PP>BR: the probability that the hypothesis is true in the light of the evidence is higher than its prior probability. From (2), this occurs if TPR>FPR, i.e. if the Likelihood Ratio is greater than 1: the evidence is more likely when the hypothesis is true than when it is false. This is not a demanding condition: evidence is confirmative if it is more accurate than a coin toss: A>0.5. Coin-toss evidence is *unconfirmative*: LR=1 and A=0.5. Unconfirmative evidence leaves probability where it was before the evidence arrived: PP=BR. Finally, evidence is *disconfirmative* if PP<BR, i.e. LR<1 and A<0.5. Disconfirmative evidence is more likely when the hypothesis is false than when it is true. Notice that, in particular, symmetric evidence is confirmative if A=TPR>0.5. Likewise, negative evidence is confirmative if NP>BR. From (2), this occurs if FNR>TNR, i.e. if the Likelihood Ratio of negative evidence is greater than 1.

Think of E as a collection of N independent pieces of evidence, $E=(E_1, E_2, \dots, E_N)$, positive or negative, each with its own Likelihood Ratio. Bayesian updating is a tug of war between confirmative and disconfirmative evidence:

$$PO = LR_1 \cdot LR_2 \cdot \dots \cdot LR_N \cdot BO \qquad (6)$$

The updating process is iterative: starting with any level of prior odds BO (except Faith, where BO is infinite or zero), confirmative evidence increases posterior odds, unconfirmative evidence leaves them unchanged, and disconfirmative evidence decreases them. The updated PO become the new BO, which is then further updated in the light of more evidence. The process is cumulative: convergence to the truth can occur by accumulation of an overwhelming amount of confirmative evidence, leading to infinite odds and PP=1, or an overwhelming amount of disconfirmative evidence, leading to zero odds and PP=0. But convergence is not assured. The tug of war does not necessarily end with a winner: the balance of evidence can leave us somewhere in the middle, where all we can say is that the hypothesis is probably true, and (one minus) probably false.

Convergence to the truth differs from Certainty. Section VI of Hume's *Enquiry Concerning Human Understanding*, entitled *Of Probability*, opens with a note on Locke:

> Mr. Locke divides all arguments into demonstrative and probable. In this view, we must say, that it is only probable that all men must die, or that the sun will rise to-morrow. But to conform our language more to common use, we ought to divide arguments into *demonstrations, proofs,* and *probabilities*. By proofs meaning such arguments from experience as leave no room for doubt or opposition.

Demonstrations are based on what we have called Faith, a prior belief in the truth or falsity of a hypothesis on the grounds of pure reason. Faith requires no evidence, and no evidence can change it. *Probabilities* are the result of the tug of war between confirmative and disconfirmative evidence, when none of the two sides manages to prevail on the other. *Proofs* occur when the tug of war has a winner. This can result from the acquisition of conclusive evidence. Multiplicative accumulation implies that even a single piece of conclusive evidence can immediately drive Posterior Odds all the way to infinity or to zero. A Smoking Gun is sufficient to prove that the hypothesis "The suspect is guilty" must be true. A Perfect Alibi is sufficient to prove that the hypothesis must be false. Or, to use another famous analogy, one black swan is sufficient to prove that the hypothesis "All swans are white" must be false.

But proofs can also result from the accumulation of overwhelming confirmative or disconfirmative evidence. Multiplicative accumulation implies that, if Likelihood Ratios are consistently confirmative (LR>1) or disconfirmative (LR<1), Posterior Odds tend to infinity or to zero. Hence, posterior probabilities converge towards Certainty, but they never reach it. We cannot *demonstrate* that all men must die, or that the sun will rise tomorrow. We can only expect it, based on an overwhelming accumulation of confirmative evidence. As they converge to one of the two boundaries of the probability spectrum, posterior probabilities cease to depend on Base Rates. In this sense, whatever the initial priors (except Faith), convergence *proves* that the hypothesis is true or false. This happens to everyone's satisfaction, leaving *no room for doubt or opposition*. However, such Certainty is not the inescapable consequence of conclusive evidence, but merely the limit of a convergent accumulation of inconclusive evidence. As such, it remains open to refutation.

For instance, the probability pair in our virus story is TPR=1, FPR=0.05, hence LR=20. Our test is a Barking Dog: if you have the virus, the test will tell you infallibly, i.e. it will never wrongly tell you that you don't have it. Hence, if it says that you don't have the virus – conclusive negative evidence: the dog didn't bark – then you certainly don't have it. But, as it turned out, the test said that you do have the virus. The test is confirmative: a Likelihood Ratio of 20 transforms Prior Odds of 0.001 into Posterior Odds of 0.02. However, since prior odds are very small (1/999), posterior odds are still small: 1/50, i.e. PP=2%.

But a 2% probability may still be worrying. What can you do to gain more comfort? You can repeat the test. Starting from the new Prior Odds of 0.02, if the test result is negative, again you are certainly safe. But if it is again positive, Posterior Odds increase to 0.4 and PP goes up to 29%. With a second positive result, you are justifiably worried. But for real panic you need a third test. Again, a negative result will put your mind completely at rest. But a positive one will increase the odds to 8 and PP to 89%. Now you are seriously freaking out. However, you can never be 100% certain: a fourth positive result would increase PP to 99.4%, a fifth to 99.97%, and so on. The updating process converges to the truth, but never reaches it. And with a Barking Dog there is always room for hope: a single negative result, no matter after how many positive ones, can still conclusively prove that you are safe. Definitely not a good reason, however, to postpone drafting your will.

One last definition: evidence is *supportive* if PP>0.5, i.e. if the probability of the hypothesis in the light of the evidence is higher than 50% or, equivalently, the Posterior Odds are greater than 1. From (4), this is true if the Likelihood Ratio is greater than the inverse of the Prior Odds. This is a much more demanding condition. For instance, our virus test, while confirmative, is far from being supportive, since its Likelihood Ratio, at 20, is much smaller than 999, the inverse of the Prior Odds. While the test is 20 times more likely to deliver a True Positive than a False Positive, the virus is 999 times more likely to be absent than to be present. Hence its odds in the light of a positive result are only 2%. In order for LR to be greater than 999, the probability of a False Positive would have to be lower than 0.1%. This proves the Humean dictum: "Extraordinary claims demand extraordinary evidence"[4]. In order to support the presence of a virus, the evidence in its favour

[4] "A wise man, therefore, proportions his belief to the evidence". Hume, Enquiries, Section X.
"The more extraordinary the event, the greater the need of its being supported by strong proofs. For those who attest it, being able to deceive or to have been deceived, these two causes are as much more probable as the reality of the event is less." Laplace, p. 17.

would have to be very strong, i.e. require a near-perfect test. Alternatively, Prior Odds would have to be higher than 0.05. As we have just seen, this would happen after a second positive test result. Only then, a third positive result would support the hypothesis that you have the virus.

Notice, finally, that symmetric evidence is supportive if A=TPR>1-BR, i.e. if the probability of error FPR is lower than the Base Rate BR.

4. The Prior Indifference Fallacy

Having analysed the general case and defined different evidence types, let's now go back to the Inverse Fallacy.

Our virus test is a Barking Dog. The doctor said: if you have the virus, the test will tell you infallibly: TPR=1. He also said that, since FPR>0, the test may wrongly tell you that you have the virus. What he didn't say was that, since the virus is rare, the probability that you have it, given a positive test result, is small: PP is only 2%. You took the test because you were especially worried about a False Negative – the nightmare scenario in which the test delivers a Miss, i.e. you think you don't have the virus but you actually have it. You wanted a test that excluded such a possibility. Attracted by this feature, you took the test and paid no attention to the frequency of the virus. As a result, you ended up believing you were close to certain death[5]. The Inverse Fallacy can open a wide gap between perceived and actual probabilities.

What causes the Inverse Fallacy? Using our notation, the fallacy consists in confusing PP with TPR. Notice that this happens in (2) if BR=0.5:

$$\mathrm{PP}=\frac{\mathrm{TPR}}{\mathrm{TPR}+\mathrm{FPR}}\approx\mathrm{TPR}$$

$$\mathrm{NP}=\frac{\mathrm{FNR}}{\mathrm{FNR}+\mathrm{TNR}}\approx\mathrm{FNR} \tag{7}$$

In (7), PP is close to TPR and NP is close to FNR because, typically, TPR+FPR is close to 1. In fact, TPR+FPR=1 exactly in case of symmetric evidence. Therefore, the Inverse Fallacy is ultimately a *Prior Indifference Fallacy*.

In our virus story, the reasoning is: I don't really know whether I have the virus or not. But I have taken this test, which the doctor says is very accurate: in fact, it is infallible at spotting the virus, and only rarely mistakes a healthy person for an infected one. If the test is really this accurate, then I am in trouble: since the test is 100% accurate at spotting the virus, and the test is saying that I have the virus, then I almost surely have it. We saw that this is a massively mistaken conclusion: the actual probability is less than 2%. The key to the blunder is in the first sentence: *I don't really know whether I have the virus or not*. Innocuous as it seems, this is equivalent to prior indifference: BR=0.5. You are implicitly assuming that, before the test, you have a 50%

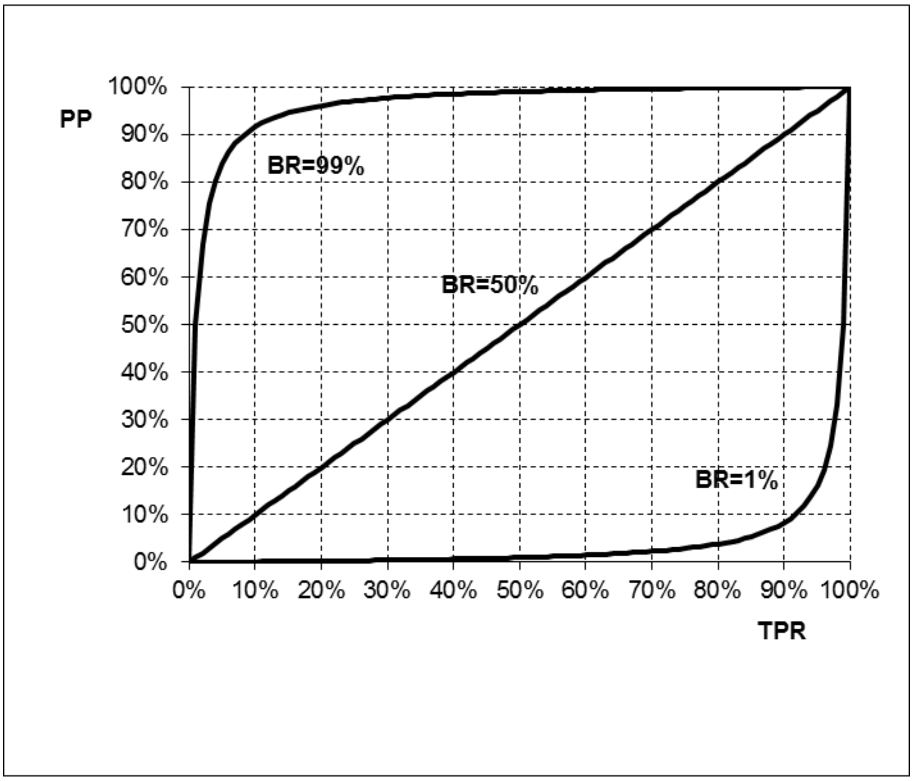
[5] Imagine the doctor had offered you a Smoking Gun, i.e. a test that, if you *didn't* have the virus, would have told you so with 100% certainty. This test would have excluded a False Positive – the unpleasant but less nightmarish scenario in which the test delivers a False Alarm, i.e. you think you have the virus but you actually don't. As you cared most about avoiding a False Negative, you wanted a Barking Dog. But, given the rarity of the virus, a Smoking Gun would have actually been a near-perfect test, with PP=1 and NP very close to 0.

chance of having the virus! Such a blatant mistake is strictly dependent on the presence of a test. In fact, imagine there was no virus test. You hear on TV that forty people have died. What is the first thing you would ask the doctor? Naturally, you would enquire about the virus frequency: how likely am I to get the virus? To your relief, the answer would be 1/1000: you have a 99.9% chance of survival. But as soon as the doctor mentions the test, you ignore the Base Rate and concentrate your attention entirely on the test's response. Moreover, as the test is very accurate – indeed perfectly accurate for what you care most: avoiding a False Negative – you take the test response as virtually infallible.

Notice from (4) that, under prior indifference, Prior Odds are equal to 1, hence Posterior Odds coincide with the Likelihood Ratio. It follows that evidence is supportive if its Likelihood Ratio is greater than 1, i.e. it is supportive if it is confirmative. In fact, PP=LR/(1+LR), which equals 0.5 if LR=1, and tends to 1 as the Likelihood Ratio increases. Under prior indifference, all it takes for evidence to be supportive is to be confirmative. The Likelihood Ratio would have to be less than 1 for PP to be lower than 0.5: only disconfirmative evidence would fail to lend support to the virus hypothesis. As long as you test positive – to any test, even a worthless one – you fall prey to the Prior Indifference Fallacy.

Figure 2 gives a graphic depiction of the relationship between PP and TPR for different levels of BR, for the particular case of symmetric evidence[6]. The relationship is positive: the higher the level of accuracy, the higher the level of support, for any given level of the Base Rate. However, the relationship is concave if BR>0.5, and increasingly so as BR tends to 1. Conversely, the relationship is convex if BR<0.5, and increasingly so as BR tends to zero. Only if BR=0.5 the relationship is 45° linear, and PP=TPR.

Figure 2 – **Relationship between posterior and anterior probabilities in a symmetric test**



[6] Notice Figure 2 is a two-dimensional representation of Figure 1.

Figure 2 makes clear that the Inverse Fallacy is due to a failure to appreciate the increasing non linearity of the relationship between accuracy and support as the Base Rate departs from the 50% indifference level. The higher the Base Rate, the larger is the underestimation of PP. The lower the Base Rate, the larger is its overestimation. In particular, a small BR (as in our virus story) implies a very convex relationship between PP and TPR, such that even a small departure from perfect accuracy implies a large drop of PP. For instance, with BR=1% (as in Figure 2) even a 1% drop from perfect accuracy (TPR=99%) implies a massive drop of PP all the way to 50%, as the probability of error equals the Base Rate. If the probability of error is higher than the Base Rate, PP falls below 50%. For instance, with TPR=95%, PP drops to 16% and with TPR=90% it drops to 8%. The Prior Indifference Fallacy hides the implications of convexity. Under prior indifference, a 1% drop in the level of accuracy translates into a 1% drop in the level of support: "Since the test is 99% accurate at spotting the virus, and the test is saying that I have the virus, then I have the virus with 99% probability". The logic is the same if the test is 95% or 90% accurate. Under prior indifference, accuracy equals support: all it takes for a symmetric test to be supportive is to be more accurate than a coin toss. In fact, imagine the doctor had said the test was only 50% accurate. This is a useless test, and the correct conclusion, following Bayes' Theorem, would be PP=BR: the probability that you have the virus after a positive response should not move from the Base Rate. But under prior indifference, a positive test result – however worthless the test – would push you towards the very wrong conclusion that your chance of survival is only 50%.

Here is the amazing paradox. Without the test, you would have reckoned you had a 0.1% probability of having the virus. With the test, after a positive response, that probability increased 20-fold: it went from 0.1% to 2%. But you thought it had gone all the way to 100%, or thereabouts. You took the test because you wanted more evidence to lead you closer to the truth, but you ended up drifting far away from it. You would have stayed much closer to the truth if you had not taken the test: you were *blinded by evidence*.

5. Hard evidence

Let's take a closer look at the evidence. Let's say the test was tried on a random sample of 20,000 individuals. We can imagine the result of the trial to be something like this:

Table 5 **Virus test trial results**

	Hypothesis is true	Hypothesis is false	TOTAL
Test is positive	20	999	1,019
Test is negative	0	18,981	18,981
TOTAL	20	19,980	20,000

Of 20,000 people, 20 of them (0.1%) had the virus. All of them tested positive, i.e. there were no False Negatives. However, there were some False Positives: of the 19,980 people who did not have the virus, 999 of them (5%) tested positive. So, of the total of 1,019 people who tested positive, 20 of them – or 1.96% – had the virus, while 98.04% did not have it. And of the total of 18,981 people who tested negative, none of them had the virus. Hence, the posterior probability of the virus, given a positive response, was

20/1019=1.96%, while the posterior probability, given a negative response, was 0/18981=0. These are precisely PP and NP, as calculated[7] in section 2.

A test is a form of evidence. In general, any sign that can be related to a hypothesis is a form of evidence about the hypothesis. Evidence can come in different shapes. A test provides *hard evidence*: It is the result of a controlled, replicable experiment, leading to the measurement of hard probabilities, grounded on empirical frequencies. But the doctor might have simply said: "I am an experienced doctor and I can spot if you have the virus. In fact, I am infallible virus spotter: give me 100 people with the virus and I will correctly identify all of them. True, I may throw out a few False Alarms, but of 100 people with no virus, I will correctly identify 95 of them. So overall I am very accurate. Do you want me to take a look at you?" You nervously consent. After performing a thorough examination, he comes back with a response: "I am sorry, but I think you have the virus. Remember: I may be wrong but – I don't think so." Again, you are very worried. The Prior Indifference Fallacy has fooled you.

In principle, the doctor's opinion could also derive from hard evidence, measured as in Table 5. In that case, results may well be similar to those in the table, with "Test is positive/negative" replaced by "Doctor says virus/no virus". With proper measurement, the resulting probabilities would be as hard as in the virus test, with a high level of accuracy translating – surprisingly, but unquestionably – into a low level of support for the virus hypothesis. Without proper measurement, however, the observation that the doctor never misses a virus overshadows the fact that False Positives are much more frequent than True Positives. As the table makes clear, what counts is not the True Positive Rate (100%) versus the False Positive Rate (5%), but the frequency of True and False Positives (20 versus 999). Neglecting this fact plays right into the hands of the Prior Indifference Fallacy. The confident doctor who invites you to rely on his accuracy makes this mistake. While not ignoring False Positives, he compares them to True Negatives (against which they appear small) rather than to True Positives (against which they are large). When he says: "Give me 100 people with the virus and I will correctly identify all of them; give me 100 people with no virus and I will correctly identify 95 of them", he is implicitly assuming an equal number of virus and no virus cases – i.e. he is falling prey to the Prior Indifference Fallacy. Indeed, from (7), under prior indifference the posterior probability is equal to the ratio of the True Positive Rate over the sum of the True and False Positive Rates. Even an honest, scrupulous expert, who correctly notates his track record, may not know what his experience means. Thus, if an accurate doctor believes you have the virus, you trust him: he is the *expert*. An expert is someone who is supposed to have tested the hypothesis many times before, and therefore has been able to catalogue his *experience* in the shape of a frequency table. His correct reasoning should be: "Since the virus is rare, the probability that this man has the virus, however confident I am that he has, is less than 2%". But thinking of False Positives as a small rate rather than a large number leads the doctor – and you – to identify accuracy with support. The failure to appreciate the nonlinear relationship between accuracy and support when the Base Rate differs from the 50% indifference level means that the expert's response can be seriously misinterpreted. Even a 50% accurate expert – a useless one, worth as much as a coin toss – may be able to produce a massive shift in probability from a low Base Rate – where the probability should stay and, without the expert response, would stay – to a grossly overestimated indifference between truth and falsity.

Like optical illusions, wrong beliefs can be impervious to the hardest evidence. That is why many people believe weird things, despite the evidence to the contrary is as hard as it gets. For example, you may have

[7] Using our notation, if N=20,000, then N·TPR·BR=20, N·FPR·(1-BR)=999 and N·TPR·BR+N·FPR·(1-BR)=1019. Hence 20/1019=PP. Notice that, by dividing the first two rows of Table 4 by the bottom Total row, we get the anterior probabilities in Table 1.

reasons to believe – as homeopaths do – that 'specially treated' water is an effective remedy against a particular disease. This is fine, as long as you can show you are right. To do so, you need to take a random sample of patients suffering from the disease, give half of them a pill soaked in your special water and the other half a sugar pill (placebo), in your chosen dosage and duration. Neither you nor the patients should know who gets which pill until the end of the treatment. At that time, count how many patients have been restored to health (however defined). Then, among them, find out how many have been treated with the special pill and how many with the placebo. For the special pill to be considered effective, you need to show that most of the recovered patients have been treated with the special pill rather than the placebo. As an example, in Table 6, 10% of patients have recovered. Of those, 1,900 have taken the special pill and 100 have taken the placebo. As a result, you can say that the probability of recovery after taking the special pill is 1900/10000=19%, much higher than the 100/10000=1% probability of recovery after taking the placebo. The special pill is highly effective.

Table 6 **Effective 'Special pill' trial results**

	Recovered	Not recovered	TOTAL
Special pill	1,900	8100	10,000
Placebo	100	9900	10,000
TOTAL	2,000	18,000	20,000

Now compare Table 6 with Table 7. The recovery rate is still 10%, but in Table 7 half of the recovered patients have been treated with the special pill and half with the placebo. As a result, the probability of recovery after taking the special pill is 10%, the same as the probability of recovery after taking the placebo – and the same as the recovery rate. The special pill is completely useless.

Table 7 **Ineffective 'Special pill' trial results**

	Recovered	Not recovered	TOTAL
Special pill	1,000	9,000	10,000
Placebo	1,000	9,000	10,000
TOTAL	2,000	18,000	20,000

Your belief about the effectiveness of the special pill should be entirely dependent on whether trial results more look like Table 6 or Table 7. And if they look like Table 7, you should abandon your belief. But homeopaths refuse to do so, despite the fact that trial results on homeopathic medicine show, repeatedly and inexorably, that its effectiveness is indistinguishable from placebo[8]. Using our notation, in Table 6 the Base Rate is 10%, the True Positive Rate – or, as it is known in clinical trials, the Sensitivity – is 95%, and the True Negative Rate – or Specificity – is 55%. Hence the Posterior Probability is 19% – almost twice the Base Rate. In Table 7, the Base Rate is also 10%, but TPR and TNR are 50%. Hence PP=10% – the same as the Base Rate.

The useless pill in Table 7 is equivalent to the useless test and the useless expert in Section 4: all worth as much as a coin toss. What keeps the homeopathic credo alive is the same phenomenon that gives credence

[8] See, for example, Goldacre (2008) and Singh, Ernst (2008).

to useless tests and worthless experts: the Prior Indifference Fallacy. Homeopaths look at the 1,000 patients in Table 7 who recovered after taking the special pill and say: "This is tough disease. Only 10% of patients recover at all. But our pill cured half of them: not bad!" The fact that the other half recovered after taking the placebo fails to dampen their enthusiasm. But the biggest boost to the homeopathic delusion comes from high recovery diseases. A common cold, for example, has, given enough time, a 100% recovery rate. This means that the probability of recovery after taking a homeopathic treatment is 100%. Enthusiastic homeopaths will gloat on this piece of hard evidence, neglecting the fact that placebo – as well as dressing up as Elvis – will have the same effect.

6. Soft evidence

If wrong beliefs can persist in spite of hard evidence, they can be utterly pervasive when hard evidence is not available. If an experiment is not possible, or if it has not been performed, the only available evidence is soft. While still based on empirical observation, *soft evidence* can only generate subjective probabilities, as determined by the observer's perception. With soft evidence, w is a *prior probability*, i.e. the observer's belief about the relative frequency of the hypothesis, while TPR and FPR measure the *perceived* accuracy of the evidence, i.e. the observer's *confidence* in using the evidence as a sign for evaluating the probability of the hypothesis.

In our virus story, the test, based on experiment, is an example of hard evidence. The doctor's opinion, based on experience, is an example of soft evidence. The doctor's accuracy has not been properly measured through a controlled, replicable experiment. Hence it is in the eye of the beholder: it is accuracy as perceived by the observer, i.e. the observer's confidence in using it as a sign for evaluating the probability that the hypothesis is true. It is the observer who decides whether the doctor's Accuracy is higher, equal, or lower than 50%, i.e. whether the doctor's opinion is confirmative, unconfirmative or disconfirmative. This is ultimately a matter of *trust*.

Most of what we believe is not the result of direct experience, but of trusting the source of the evidence. That's how we know, for instance, that the Coliseum is in the centre of Rome, even if we have never been there. The accuracy we attach to soft evidence is ultimately our decision. We decide to trust the Rome Tourist Guide as to the Coliseum's whereabouts, because we attach to it a very high TPR and a very low FPR – i.e. a very large Likelihood Ratio. The guide has our full trust: we regard its indication as conclusive evidence. If the guide says it, it must be true. It is an entirely reasonable decision, based on attaching a zero probability to the chance that the guide's authors may have made a mistake or lied.

We do the same with all evidence. To each piece we attach a Likelihood Ratio, which is ultimately based on trust. The evidence may come from myriad of sources: newspapers, TV, books, conversations; teachers, parents, friends; scientists, politicians, clerics. The trust we place on evidence is greatly influenced by its source. Smoking is bad for your health is more likely to be true if said by your doctor than by your mother; IBM shares are more likely to be a good investment if you hear it from Warren Buffett than from your uncle. Greed is more likely to be a capital sin if the Bible says so than if you read it on Hello! Magazine. In fact, you may go as far as giving the doctor, Buffett and the Bible your full trust. If they say it, it must be true. As with the Rome guidebook, it is your decision.

Trust does not require Certainty. As long as it is highly trusted, the source of evidence will have a large influence on our beliefs. Of course, trust may not be omnipotent: if Buffett says that elephants fly, we won't believe him (although I know a few fans who would give it a thought). This is because we attach a miniscule prior probability to flying elephants: whatever anyone says, we won't believe it until we see one. But as long as the source of evidence does not stray too wide from the confines of its credibility, prior indifference kicks in: unfiltered by our priors, evidence blinds us. If BR=50%, hence BO=1, then PO=LR: support equals accuracy. As a result, our posterior probability is *entirely* determined by trust. Under prior indifference, *what* we believe depends on *who* says it.

The Prior Indifference Fallacy explains the power of experts. A confident doctor says you have the virus. His assessment may be perfectly honest: based on years of experience, he believes TPR=1 and FPR=0.05. What he is missing is the fact that, since the virus is rare, the number of False Positives is much larger than the number of True Positives, despite a low False Positive *Rate*. Focus on a small rate of False Positives rather than on their large number leads the doctor – and you – to confuse accuracy with support. As a result, a posterior probability of less than 2% appears as high as 95%.

The Prior Indifference Fallacy gives experts an incentive to be overconfident. The doctor in our story is honest and may well be right, i.e. his confidence may reflect his true accuracy. But other experts may not be as scrupulous. An easy way to increase confidence is to increase y. Since the focus is on the hypothesis, it is important not to miss it whenever it is true. In our virus story, you care most about avoiding False Negatives. In the extreme case, the doctor could ensure TPR=1 by telling all his patients that they are infected. Of course, this approach would imply FPR=1 and, from Bayes' Theorem, PP=BR: the doctor's assessment would be obviously worthless. Nonetheless, under prior indifference, he would be able to gain a totally undeserved 50% support.

If this seems a bit stretched, imagine the hypothesis is "There will soon be a stock market downturn". What can an expert – let's call him Dr. Doom – do in order to gain support? He can call a downturn as often as possible. By doing so, he will maximise the chance of spotting all or most downturns. Clearly, there will be many times when his warning will not come true: the trade-off between Type I and Type II errors implies that an increase in TPR can only come at the cost of a higher FPR. But, as long as the public is worried about downturns, False Alarms will likely be forgiven and soon forgotten, and Dr. Doom will be hailed as an oracle.

An alternative to increasing TPR would be to decrease FPR. Ideally, a Smoking Gun (FPR=0) would be preferable to a Barking Dog (FNR=0), as it would imply PP=1 irrespective of BR and TPR. A Smoking Gun does not need prior indifference: it is prior-free, conclusive evidence. However, the cost of such infallibility would have to be a lower TPR, i.e. a higher False Negative Rate: the expert would have to refrain from calling market downturns too often, thereby incurring into many False Negatives. But False Negatives are much worse than False Positives. A worried public will forgive False Alarms, but will penalise Misses. Increasing TPR is therefore a better trick. Inflated focus on TPR, coupled with dimmed attention to FPR, is a form of Confirmation Bias. Unscrupulous experts can exploit the bias by emphasising good calls and obfuscating bad calls. If nobody keeps the score, what counts is what is remembered. A high TPR and a hidden FPR imply a high posterior probability: if Dr. Doom calls it, the public believes there is a high probability of a market downturn.

By keeping the limelight on the high TPR and obfuscating the resulting high FPR, the Confirmation Bias gives unscrupulous experts an incentive to be overconfident. *Overconfidence* is the difference between an artificially high TPR and the true TPR that would result from an honest prediction effort. It pays to be overconfident if the gain from a higher TPR exceeds the loss from a higher FPR.

Unscrupulous experts have an obvious disadvantage: their trick can be easily spotted. After a few False Alarms, their credibility will rapidly fade. But Dr. Doom has a whole bag of tricks up his sleeve. When his call for a market downturn turns out to be false, he can always push it forwards and, when a downturn finally arrives, vindicate his prediction. He can also appeal to prudence: it is “better safe than sorry”. And he can cultivate his credibility by trumpeting True Positives with fanfare and quietly brushing False Positives under the carpet.

7. Perfect Ignorance

Prior Indifference (BR=0.5) is on the opposite side of Faith (BR=1 or BR=0). It is perfect ignorance: no clue at all about whether the hypothesis under investigation is true or false.

Imagine an urn containing 100 balls, black and white, in unknown proportions. What is the probability of extracting a white ball? The immediate answer is: no idea, we just don’t know. This feeling of helplessness is what is known as Knightian uncertainty[9]. We would rather not answer the question but, if forced to, our thinking may be: there are 99 equiprobable proportions, ranging from 1 white/99 black to 99 white/1 black. Hence we take their average: 50%. Under the circumstances, it is clearly the best answer. It is the same answer we would give if we knew that the balls were 50% white and 50% black. But under Knightian uncertainty we don’t know the actual proportion – in fact we know that it is almost surely different from 50/50. It is precisely such ignorance that motivates our answer.

Despite the equivalence, if we had to choose between betting on the extraction of a white ball from an urn with a known 50/50 proportion and an urn with an unknown proportion, we would prefer the former. This is known as Ellsberg paradox, or ambiguity aversion[10]. We prefer known risk to unknown uncertainty. But Prior Indifference is the starting point of both.

So BR=0.5 does not necessarily mean that we know the prior probability of the hypothesis is 50%. It may simply mean that we know nothing at all – nothing that allows us to differentiate between true and false: Perfect Ignorance. Do I have the virus? If your answer to this question is: I have no idea, you are in the grip of the Prior Indifference (or Perfect Ignorance) Fallacy.

Why is it a fallacy? Because it is hardly ever true that we have no idea. Most times our priors already contain plenty of background evidence that we wrongly ignore. As ex US Secretary of Defense Donald Rumsfeld famously said[11]:

> There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don’t know we don’t know.

But there is fourth element in Rumsfeld’s matrix:

Table 8 **Rumsfeld’s Matrix**

	Known	Unknown
Aware	Known knowns	Known unknowns
Unaware	Unknown knowns	Unknown unknowns

[9] Knight (1921).

[10] Ellsberg (1961), Fox, Tversky (1991).

[11] www.defense.gov/transcripts/transcript.aspx?transcriptid=2636

Unknown knowns are things that we are not aware we know. It is available evidence that we fail to take into account because a blind spot prevents us from seeing it. Prior indifference renders the Base Rate an unknown known.

You hear on television that forty people have recently died from a lethal virus. You are not a Martian catapulted on earth with no knowledge of earthly matters: although you are worried, you can easily find out that the virus is rare: it hits about one in a thousand.

But hang on. 1/1000 is the probability of extracting an infected person from the general population. This is not what you are after: you want to know the probability that *you* have the virus. This could be properly assessed only by comparing yourself to others who are more *like* you: people who share the same, or at least a comparable probability of getting infected. But what does *comparable* mean? For instance, future genetic research may reveal a link between the virus and a particular gene, which is found in, say, only 2% of the population. If a person does not carry that gene, he will certainly not get the virus. But if he has it, the probability of getting it is 5% (apologies to biologists: it is an exemplification). Thanks to this discovery, we would find that the 1/1000 population Base Rate is really the product of 2% times 5%. So, in a sample of 20,000 people, 20 carry the gene and will get the virus, 380 carry the gene but will not get it, and the rest have no gene and therefore no virus. Or perhaps the gene is present in only 1% of the population, and those who carry it have a 10% probability of getting the virus. Or maybe it is such a rare gene that has a 0.5% frequency, and the unlucky ones have a 20% chance of being affected. And why not go all the way: only one in a thousand have the gene and are therefore predestined to certain death.

The Base Rate and, with it, the posterior probability of the hypothesis depend on the definition of the relevant population. You don't try the virus test on people who cannot get the virus, just as you don't try hair conditioner on bald men. What is the appropriate Base Rate? Given the current state of knowledge, it is 1/1000. But it could be completely different, depending on the definition of the appropriate reference class[12]. We can think of the reference class as an image of the state of knowledge about the virus. The more we know, the smaller the reference class. Indeed, knowledge can be defined as a progressive narrowing down of possibilities. In Sherlock Holmes' immortal words: "When you have eliminated the impossible, whatever remains, however improbable, must be the truth"[13]. The smaller the reference class, the higher the Base Rate for individuals belonging to that class. In the limit, knowledge about the virus could become as complete as to allow us to narrow down the population to precisely *those* one-in-a-thousand individuals who will certainly get the virus. Getting the virus would then be either a certainty or an impossibility.

This uncertainty about the appropriate reference class is distinctly Knightian. Given current knowledge, the Base Rate is 1/1000, but with increased knowledge it could be anywhere between 0 and 1 – like extracting from an urn with white and black balls in unknown proportions. It is in this state of uncertainty, and with the aim of increasing your knowledge, that you asks the expert doctor: What is the chance that I will get the virus? Remember the doctor is very accurate: he is infallible at spotting the virus and mistakes a healthy individual for an infected one only 5% of the times. After the doctor's response, you no longer see yourself as a comparable member of the general population. The 1/1000 Base Rate – so clear and consequential until then – is driven to the background: it becomes an unknown known. You no longer know which reference class you belong to, hence you cannot define the relevant Base Rate. And since an undefinable Base Rate

[12] Hajek (2007).

[13] Conan Doyle, Chapter 6.

could be anywhere between 0 and 1, you pick the neutral midpoint: you become prior indifferent. You simply think: I may or may not have the virus, attach an equal chance to the two possibilities, and let the doctor decide. And if the doctor says you have the virus, you believe him. The urge to resolve this uncomfortable state of Knightian uncertainty is what consign you into the hands of the expert. Under perfect ignorance, you replace support with accuracy, confidence and, ultimately, trust.

Seen in this light, prior indifference is a distortion of Bayesian updating. While a correct update takes BO as given and increases or decreases it according to the Likelihood Ratio of new evidence, prior indifference triggers an inadvertent shift of the Base Rate to 50% *before* the update takes place. As a result, the update builds on Knightian uncertainty and perfect ignorance, rather than on prior beliefs.

8. Prior Indifference and other fallacies

The Inverse Fallacy is often illustrated using Tversky and Kahneman's cab problem[14]:

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> 85% of the cabs in the city are green and 15% are blue.
>
> A witness identified the cab as blue. The Court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colours 80% of the time and failed 20% of the time.
>
> What is the probability that the cab involved in the accident was blue rather than green?

The common answer is 80%: go along with the witness. As in our virus story, the "expert" rules the day. But it is the wrong answer. The true probability is half of that. Let's see:

Hypothesis: The cab involved in the accident is blue. Evidence: A witness says so.

1. What is the prior probability that the cab is blue? 15% of the cabs are blue: BR=15%.
2. What is the probability that the witness says the cab is blue, if indeed it is blue? The Court says TPR=80%.
3. What is the probability that the witness says the cab is blue, if it is actually green? The Court says FPR=20%.

This is a case of symmetric evidence, with an equal probability of Misses and False Alarms. Hence PP=41%. The mistake is due to the Inverse Fallacy, which is ultimately a Prior Indifference Fallacy. Under prior indifference, PP=TPR=80%. What causes prior indifference? Why is it so immediately powerful? The answer can be found by contrasting the original cab problem with a slightly modified version, where the base information is changed to:

1a. Green cabs are involved in 85% of the accidents.

The modified version is formally identical to the original: the prior probability that the cab is blue is still 15%. Had there been no witness to the accident, 15% would have been the obvious answer in both cases. But,

[14] Tversky, Kahneman (1982), Kahneman (2011), Chapter 16.

after the witness testimony, the common answer in the modified version is much lower than 80% and close to the true 41% posterior probability.

Why is the witness testimony much less influential in the modified version? It is because 1a is not merely a *statistical* Base Rate: it is a *causal* Base Rate. 1a gives us a *reason* to believe that blue cabs are less likely to be involved in the accident. In 1a we may not even know the proportion of green and blue cabs, but we know that green cabs are much more accident-prone than blue cabs. So when the witness tells us that the cab was blue we see the need to balance this piece of information with the fact that green cabs are run by sloppy drivers.

Statistical Base Rates are not beliefs. Hence they are ignored: they are the unknown knowns that give power to experts, whether they are accurate – like the accident witness and the doctor in our virus story – or merely confident – like Dr. Doom and other unscrupulous forecasters. Under prior indifference, we are blinded by evidence, even when it is perfectly useless. Causal Base Rates, on the other hand, are beliefs. Hence they are *not* ignored, but are modified by evidence according to Bayes' Theorem. Causal Base Rates prevent prior indifference and therefore, if correct, keep us closer to the true posterior probability.

Notice that nothing substantial would change if we witness the accident ourselves, and are 80% sure that the cab was blue. Despite our confidence, we should account for the fact that the Base Rate favours green cabs and adjust our prediction accordingly. Like the expert doctor in section 5, our reasoning should be: Since there are many more green cabs than blue cabs, the probability that the cab was blue must be adjusted downwards, however confident I am that it was indeed blue[15].

In fact, Base Rate neglect is commonly referred to as *representativeness*, defined as a probability judgement based on the similarity between the evidence and the object under investigation, where we are the 'experts' evaluating the evidence. Evidence can come from a simple description, such as:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
>
> Question: Is Linda:
>
> 1. A bank employee
> 2. A Greenpeace supporter.

This is a slightly modified version of another well-known Kahneman and Tversky experiment[16]. The description is not accurate enough to answer the question with certainty. So we have to go with the most probable choice: is Linda more likely to be a bank employee or a Greenpeace supporter? Let's see:

Hypothesis: Linda is a bank employee/Greenpeace supporter. Evidence: Linda's description (let's call it E).

The problem is best looked at in odds form:

$$\frac{\mathrm{PO}_2}{\mathrm{PO}_1} = \frac{\mathrm{LR}_2}{\mathrm{LR}_1} \cdot \frac{\mathrm{BO}_2}{\mathrm{BO}_1} \qquad (8)$$

[15] Unless, that is, I am absolutely certain: TPR=1, in which case PP=1, irrespective of the Base Rate.

[16] Tversky, Kahneman (1984), Kahneman (2011), Chapter 15.

where 1 is 'bank employee' and 2 is 'Greenpeace supporter'. Let's call $K=BO_2/BO_1$ the ratio of the prior odds of Greenpeace supporters and bank employees. Without a description, Linda would clearly be 1/K times more likely to be a bank employee than a Greenpeace supporter: $PO_1=PO_2/K$. How accurate a portrait of a bank employee/Greenpeace supporter is E? Again, it is not easy to say in absolute terms, but surely Linda *looks* much more like a Greenpeace supporter than a bank employee: $LR_2>LR_1$. Let's also say that E is totally unconfirmative as a description of a bank employee: $LR_1=1$. Finally, for simplicity let's assume symmetry, so that accuracy A=TPR and LR=A/(1-A). Then (8) becomes:

$$\frac{PO_2}{PO_1}=\frac{A_2}{1-A_2}\cdot K \tag{9}$$

Therefore, given E, the odds that Linda is a Greenpeace supporter are greater than the odds that she is a bank employee if $A_2>1/(1+K)$.

For example, if K=10%, then Linda is more likely to be a Greenpeace supporter if $A_2>91\%$. If K=1% the required A_2 is 99% and if K=20% it is 83%. In any case, the required level of accuracy is very high. The accuracy of a Greenpeace supporter description can go from 0 ("Linda is an avid game hunter and ivory collector") to 1 ("Linda is the captain of the Rainbow Warrior"), passing through the totally unconfirmative 0.5 ("Linda is blonde and likes chocolate"). E is plausibly more than 50% accurate as a description of a Greenpeace supporter, but it is unlikely to be as high as 80%. Hence the right conclusion, according to Bayes' Theorem, is that Linda is more likely to be a bank employee than a Greenpeace supporter.

But this is not what most people think. The most common answer is that, given E, Linda is more likely to be a Greenpeace supporter. The reason, once again, is the Prior Indifference Fallacy. Under prior indifference, K=1, hence the required A_2 falls down to 50%: Linda is more likely to be a Greenpeace supporter than a bank employee if she is simply more likely than not to be a Greenpeace supporter.

Consider now a slight variation. Question: Given description E, is Linda:

1. A bank employee
2. A bank employee who is also a Greenpeace supporter.

The problem is essentially the same. Again K<1, this time not only statistically but logically: 2 *must* be a subset of 1. Also, $LR_2>1$: Linda looks more like a bank employee *and* Greenpeace supporter than like a simple bank employee. Hence we can draw the same conclusion: according to Bayes' Theorem, Linda is more likely to be a bank employee, *unless* E is a very accurate description of a bank employee *cum* Greenpeace supporter.

Again, experimental evidence shows that most people think 2 is more likely than 1. Kahneman and Tversky call it the *Conjunction Fallacy*[17], referring to the impossibility that K>1 and implying that, therefore, PO_1 *must* be bigger than PO_2. However, as we have seen, that is not necessarily the case: there can be sufficiently accurate descriptions of Linda, such that it *is* rational to conclude that 2 is more likely than 1, despite a lower Base Rate (for example: "Linda is a bond trader who devotes her entire annual bonus to environmental causes").

Linda is judged to be more likely a Greenpeace supporter than a bank employee because her description is more representative of the former than of the latter. In simpler words, Linda *looks* more like a typical

[17] Tversky, Kahneman (1984).

Greenpeace supporter than a typical bank employee. Such evidence obfuscates the prevalence of bank employees over Greenpeace supporters in the general population which, in the absence of a description, would naturally imply the opposite probability ranking.

I call this prior indifference because it gets to the crux of the matter: the Inverse Fallacy. People confuse the probability of the hypothesis, given the evidence, with the probability of the evidence, given the hypothesis. And they do so because they assume the hypothesis is equally likely to be true or false.

Prior indifference also explains probability judgements in response to neutral, unconfirmative evidence. For instance, faced with a totally unrepresentative description of Linda (e.g. "Linda is blonde and likes chocolate"), the right conclusion, according to Bayes' Theorem, would be to stick to the Base Rate. LR=1 implies PO=BO: neutral evidence is the same as no evidence. But this is not what happens empirically. Given an irrelevant description, people tend to assign the same probability to Linda being a bank employee or a Greenpeace supporter, just as they assign 50% support to the predictions of a useless coin-tossing expert. They are prey to the Prior Indifference Fallacy.

Prior indifference underlies another well-documented cognitive heuristics, known as *anchoring*.

One of the experiments used to illustrate anchoring involved two groups of visitors at the San Francisco Exploratorium[18]. Members of the first group were asked:

> Is the height of the tallest redwood more or less than 1,200 feet?

while members of the second group were asked:

> Is the height of the tallest redwood more or less than 180 feet?

Subsequently, members of both groups were asked the same question:

> What is your best guess about the height of the tallest redwood?

As it turned out, the mean estimate was 844 feet for the first group and 282 feet for the second group. People were anchored to the value specified in the first, priming question. The anchoring index was (844-282)/(1200-180)=55%, roughly in the middle between no anchoring and full anchoring. This index level is typical of other similar experiments.

Why is judgement influenced by irrelevant information? It is for the same reason why, in Linda's experiment, an unconfirmative description is not equivalent to no description. Evidence can blind us not only when it is relevant and purposefully sought, but also when it is irrelevant and incidentally assimilated. Among visitors, there will be people who have quite a good sense of the height of the tallest redwood (it is called Hyperion and it is 379 feet high), some people who have only a vague sense and some who have no idea. The less one knows about redwoods, the closer he is to the state of perfect ignorance that characterizes prior indifference. Under perfect ignorance, the number in the priming question acts as a neutral reference point, around which the probability that the tallest redwood is higher/shorter is deemed to be 50/50. Asked to give a number, people with little or no knowledge of redwoods will choose one around the reference point, thus skewing the group average towards it.

[18] Jacowitz, Kahneman (1995), Kahneman (2011), Chapter 11.

In the redwoods experiment the priming question may be thought to contain a modicum of information – uninformed people may take the number as an indication of the average height of redwoods. But anchoring works even when priming information is utterly and unequivocally insignificant. In another experiment, a wheel of fortune with numbers from 0 to 100 was rigged to stop only at 10 or 65. Participants were asked to spin the wheel and annotate their number, and then were asked:

> What is your best guess of the percentage of African nations in the United Nations?

The average answer of those who saw 10 was 25%, while the average of those who saw 65 was 45%. Prior indifference is seen here in its clearest and most disturbing capacity.

We crave for and absorb information without necessarily being aware of it. Bayesian updates on unconfirmative evidence should be inconsequential: LR=1. But inconsequential evidence may influence our thoughts, estimates, choices and decisions much more than we would like to think. To protect against such danger, we should not only try to focus on relevant evidence, but also actively shield ourselves against irrelevant evidence – an increasingly arduous task in our age of information superabundance.

Another prominent cognitive heuristic is *availability*[19]. The availability heuristics is the process of judging frequency based of the ease with which instances come to mind. The area in which availability has been most extensively studied is risk perception[20].

As an example, let's take aviophobia. When someone is terrified of flying, there is no point telling him that airplanes are safer than cars. The safest means of transportation – is the typical reply – is a car *driven by me*. This illusion of control is caused by an obviously improper comparison between innumerable memories of safe car driving and many vivid episodes of catastrophic plane crashes.

Like representativeness and anchoring, availability is a probability update in the light of new evidence. But with availability evidence comes from within: our own memory. Far from being a passive and faithful repository of objective reality, memory is a highly reconstructive process, heavily influenced by feelings and emotions. As we try to assess the relative odds of a fatal plane accident versus a fatal car accident, we may well be aware that airplane crashes are more infrequent than car crashes. But when we update Base Rates by retrieving evidence from memory, we find that instances of plane crashes are more easily available than instances of car crashes.

This is essentially equivalent to Linda's problem. Here we have BO_1=Prior Odds of fatal car accidents and BO_2=Prior Odds of fatal airplane accidents, with $BO_1>BO_2$: car travel is statistically riskier than air travel. Evidence consist of retrieved memory. Let's again assume symmetry, hence accuracy A=TPR. Just as Linda's description can be a more or less accurate portrayal of a Greenpeace supporter or a bank employee, the availability of instances of airplane or car accidents defines the accuracy of our memory. Again mirroring Linda's example, let's assume $LR_1=1$: memory is neutral with respect to car accidents. A_2, the availability of fatal airplane accidents, is higher than A_1. But how much higher should it be, for air travel to be *perceived* as riskier than car travel? Again, the limit is given by (9), where K is the relative riskiness of air travel versus car travel. If air travel were as risky as car travel (K=1), all that would be necessary for airplanes to be perceived as riskier than cars would be more than neutrally available memories of airplane accidents: $A_2>50\%$. But for lower values of K the required A_2 is higher. For instance, if K=10% (as seems to be the case in the US)[21], A_2

[19] Tversky, Kahneman (1973), Kahneman (2011), Chapter 12 and 13.

[20] Slovic (2000).

[21] http://en.wikipedia.org/wiki/Transportation_safety_in_the_United_States.

needs to be higher than 90% – which may be the reason why aviophobia is confined to a minority of exceedingly impressionable types (such as, apparently, Joseph Stalin).

But what if aviophobes are right? Aviation safety is usually defined in terms of deaths per kilometre. This answers the question: if I travel from London to Edinburgh, am I safer going by plane or by car? The answer is crystal clear: airplanes win hands down. Similarly if safety is measured in deaths per hour. Given the same journey, measured in either distance or time, planes are much safer than cars. However, these two measures hide the fact that most airplane accidents happen at takeoff or landing, which occupy only a small percentage of journey distance and time. A different question is: what is the probability of dying in an airplane journey versus a car journey? When safety is measured in terms of deaths per journey, the answer seems to be unequivocally the opposite: car journeys are safer. This may be the measure in the back of our mind each time we board a plane. And while few of us go to extremes, we are ever so slightly more anxious when we are on a plane than when we are driving a car. Since we are not sure how to define the appropriate reference class for transportation safety, we tend to neglect Base Rate differences: K=1. And, as airplane accidents are more available than car accidents, such prior indifference explains our discomfort.

Concluding remarks

The idea that the accumulation of evidence leads to the truth is a powerful engine of progress. People may start from different priors but, as long as they look at the same evidence, they should, and normally do converge to the same truths. Right from the start[22], we are natural Bayesians, innately predisposed to learn about the world through observation and experience. As a result, in the big sweep of history and despite casual appearance to the contrary, humanity was witnessed a secular decline in the amount of nonsense people believe in.

Yet, if Bayesian updating worked perfectly, the world would be a different place – not necessarily better, perhaps, but surely not one still fraught with illusions, faulty reasoning and wrong beliefs[23]. Combating these flaws requires a clear understanding of where and why Bayesian updating gets its cramps.

The Inverse Fallacy is a distortion of Bayesian updating. Different terms have been used to denote it. Among others: Invalid inversion, Error of the transposed conditional, Base Rate fallacy or neglect, Prosecutor's or Juror's fallacy[24]. But we have claimed here that the best way to think about it is to call it what it ultimately is: a Prior Indifference Fallacy. Prior indifference is closely related to Base Rate neglect: being indifferent about whether a hypothesis is true or false implies ignoring its Base Rate. But the crucial attribute of faulty thinking is not inattention or neglect of evidence. Like optical illusions, prior indifference persists despite our full attention. It is there not because we ignore evidence, but because we are *blinded* by it.

The Prior Indifference Fallacy should not be seen as a systematic flaw or an automatic reflex. People are not dumb – Bayesian updating works well in most circumstances. But when it doesn't, the phenomenon cannot be simply dismissed as a casualty of semantic confusion or ineffective communication, vanishing once it is made transparent through a more explicit description[25]. Again, as with optical illusions, we can and do understand that we are making a mistake. But the illusion does not go away once we understand it. This is

[22] Gopnik (2009).

[23] See for example Shermer (2002, 2011) and Law (2011).

[24] Bar-Hillel (1980), Thompson, Schumann (1987), Koehler (1996), Villejoubert, Mandel (2002), Senn (2013).

[25] As claimed in Cosmides, Tooby (1996), Koehler (1996).

what makes prior indifference particularly insidious. Even hard evidence does not make us immune. But it is with soft evidence that the effects of prior indifference can be most pervasive. Under prior indifference, support equal accuracy. And, with soft evidence, accuracy equals confidence, and ultimately trust.

Prior indifference empowers experts and gives them an incentive to be overconfident. Thus, as long as we trust the source of the evidence, even a useless expert, worth as much as a coin toss, can produce a potentially large shift in our probability estimates. And unscrupulous experts – including ourselves – can manipulate our trust by artificially boosting their True Positive Rate and hiding the consequent increase in their False Positive Rate.

But trust is a double-edged sword. If high trust can give credence to worthless experts, thereby moving people to accept hypotheses to which they would otherwise assign low priors, low trust can have the exact opposite effect: it can induce people to reject hypotheses that have high priors. This is the ultimate reason why people believe weird things: it is not that they ignore the evidence, but they distrust it[26]. Distrust of evidence is why otherwise rational and knowledgeable people believe that lunar landings were fakes, that some secret powers killed JFK and destroyed the Twin Towers, that "alternative" medicine works, and hundreds of other follies.

What should we do to avoid prior indifference? We should resist the sirens of Knightian uncertainty and properly place new evidence within the confines of what we already know. Correct priors guard us against Perfect Ignorance, keep us closer to the truth and prevent us from getting blinded by evidence. Of course, correct priors are just a good starting point. They are neither a necessary nor a sufficient condition for convergence to the truth. Unless we can find conclusive evidence, convergence can only occur as a result of a thorough tug of war between confirmative and disconfirmative evidence, making sure that we gather plenty of it on both sides of the rope.

Priors must be constantly updated, but should never be ignored. Or – as reprised by many but, it seems, first expressed by the New York Times editor Arthur Hays Sulzberger: It is good to keep an open mind, but not so open that your brain falls out.

[26] Matthews (2005) makes the same point. But what he calls "hard facts" are not the same as hard evidence. They are soft evidence, which can be distrusted by otherwise rational people. Hard evidence is much more difficult to distrust – although, as with homeopathy, it can be done!

Bibliography

M. Bar-Hillel (1980), The Base Rate Fallacy in Probability Judgments, Acta Psychologica, 44, 211-233.

W. Casscells, A. Schoenberger, T. Greyboys (1978), Interpretation by Physicians of Clinical Laboratory Results, New England Journal of Medicine, 299, 999-1000.

A. Conan Doyle, The Sign of Four, in The Penguin Complete Sherlock Holmes, Penguin.

L. Cosmides, J. Tooby (1996), Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment and Uncertainty, Cognition, 58(1), 1-73.

D. Ellsberg (1961), Risk, Ambiguity, and the Savage Axiom, Quarterly Journal of Economics, 75, 643-669.

C. Fox, A. Tversky (1995), Ambiguity Aversion and Comparative Ignorance, Quarterly Journal of Economics, 110, 3, 585-603. In Kahneman, Tversky, Eds. (2000), 30.

T. Gilovich, D. Griffin, D. Kahneman, Eds. (2002), Heuristics and Biases. The Psychology of Intuitive Judgment, Cambridge University Press.

B. Goldacre (2008), Bad Science, Harper Collins.

A. Gopnik (2009), The Philosophical Baby, Random House.

A. Hajek (2007), The Reference Class Problem is Your Problem Too, Synthèse, 156, 185-215.

D. Hume, Enquiries Concerning Human Understanding, Clarendon Press, Oxford.

K.E. Jacowitz, D. Kahneman (1995), Measures of Anchoring in Estimation Tasks, Personality and Social Psychology Bulletin, 21, 1161-1166.

D. Kahneman, P. Slovic, A. Tversky, Eds. (1982), Judgment under Uncertainty: Heuristics and Biases, Cambridge University Press.

D. Kahneman, A. Tversky (1973), On the Psychology of Prediction, in Kahneman, Slovic, Tversky, Eds. (1982), 4.

D. Kahneman, A. Tversky, Eds. (2000), Choices, Values, and Frames, Cambridge University Press.

D. Kahneman (2011), Thinking, Fast and Slow, Allen Lane.

F.H. Knight (1921), Risk, Uncertainty and Profit, BeardBooks.

J.J. Koehler (1996), The Base Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges, Behavioral and Brain Sciences, 19, 1-17.

P. Laplace, A Philosophical Essay on Probabilities, Merchant Books.

S. Law (2011), Believing Bullshit, Prometheus Books.

R. Matthews (2005), Why do People Believe Weird Things?, Significance, 2, 182-184.

S. Senn (2013), Invalid Inversion, Significance, 10, 40-42.

M. Shermer (2002), Why People Believe Weird Things, Henry Holt and Co.

M. Shermer (2011), The Believing Brain, Henry Holt and Co.

S. Singh, E. Ernst (2008), Trick or Treatment? Alternative Medicine on Trial, Random House.

P. Slovic (Ed.) (2000), The Perception of Risk, Earthscan Publications.

W.C. Thompson, E.L. Schumann (1987), Interpretation of Statistical Evidence in Criminal Trials – The Prosecutor's Fallacy and the Defense Attorney's Fallacy, Law and Human behaviour, 11, 167-187.

A. Tversky, D. Kahneman (1973), Availability: A Heuristic for Judging Frequency and Probability, in Kahneman, Slovic, Tversky, Eds. (1982), 8.

A. Tversky, D. Kahneman (1980), Causal Schemas in Judgments under Uncertainty, in Kahneman, Slovic, Tversky, Eds. (1982), 11.

A. Tversky, D. Kahneman (1982), Evidential Impact of Base Rates, in Kahneman, Slovic, Tversky, Eds. (1982), 10.

A. Tversky, D. Kahneman (1984), Extensional versus Intuitive Reasoning: The Conjunction Fallacy, in Probability Judgment, in Gilovich, Griffin, Kahneman, Eds. (2002), 1.

G. Villejoubert, D.R. Mandel (2002), The Inverse Fallacy: An Account of Deviations from Bayes's Theorem and the Additivity Principle, Memory and Cognition, 30 (2), 171-178.