

Sometimes the advice from parents is specifically aimed against credulity towards other adults in the population. The following is a piece of advice that parents need to give their children: 'If any adult asks you to come with him and says that he is a friend of your parents, don't believe him, however nice he seems and even (or especially) if he offers you sweets. Only go with an adult that you and your parents already know, or who is wearing a policeman's uniform.' (A charming story recently appeared in the English newspapers in which Queen Elizabeth the Queen Mother, aged 97, told her chauffeur to stop the car when she noticed a crying child who was apparently lost. The kind old lady got out to comfort the little girl and offered to take her home. 'I can't,' wailed the child, 'I'm not allowed to talk to strangers.') A child is called upon to exercise the exact opposite of credulity in some circumstances: a tenacious persistence in believing an earlier adult statement in the face of what may be a tempting and plausible – but contradictory – later statement.

On their own, then, the words 'gullible' and 'credulous' are not quite right for children. Truly credulous people believe whatever they have most recently been told, even if this contradicts what others have told them before. The quality of childhood that I am trying to pin down is not pure gullibility but a complex combination of gullibility coupled with its opposite – stubborn persistence in a belief, once acquired. The full recipe, then, is extreme early gullibility followed by equally obstinate subsequent unshakeability. You can see what a devastating combination this could be. Those old Jesuits knew what they were about: 'Give me the child for his first seven years, and I'll give you the man.'

7

UNWEAVING THE UNCANNY

*... though no great minist'ring reason sorts
Out the dark mysteries of human souls
To clear conceiving...*

JOHN KEATS, 'Sleep and Poetry' (1817)

The eminent fertility specialist Robert Winston imagines the following advertisement, placed in the newspaper by an unscrupulous quack doctor, aimed at people who want their next baby to be, say, a son (the sexism underlying this assumption is not mine but could be found unquestioned all over the ancient world, and still in many places today). 'Send £500 for my patient recipe to make your baby a boy. Money refunded in full if I fail.' The money back guarantee is intended to establish confidence in the method. In fact, of course, since boys turn up anyway on approximately 50 per cent of occasions, the scheme would be a nice little earner. Indeed, the quack could safely offer compensation of, say, £250 for every girl born, over and above the money back guarantee. He would still show a tidy profit in the long run.

I used a similar illustration in one of my Royal Institution Christmas Lectures in 1991. I said I had reason to believe that among my audience was a psychic, clairvoyant individual, capable of influencing events purely by the power of thought. I would try to flush this individual out. 'Let's first establish,' I said, 'whether the psychic is in the left half or the right half of the lecture hall.' I invited everybody to stand up while my assistant tossed a coin.

Everybody on the left of the hall was asked to 'will' the coin to come down heads. Everybody on the right had to will it to be tails. Obviously one side had to lose, and they were asked to sit down. Then those that remained were divided into two, with half 'willing' heads and the other half tails. Again the losers sat down. And so on by successive halvings until, inevitably, after seven or eight tosses, one individual was left standing. 'A big round of applause for our psychic.' He must be psychic, mustn't he, because he successfully influenced the coin eight times in a row?

If the lectures had been televised live, instead of recorded and broadcast later, the demonstration would have been much more impressive. I'd have asked everybody who watched it whose surname begins before J in the alphabet to 'will' heads and the rest tails. Whichever half turned out to contain the 'psychic' would have been divided in half again, and so on. I'd have asked everybody to keep a written record of the order of their 'willings'. With two million viewers, it would have taken about 21 steps to narrow down to a single individual. To be on the safe side I'd have stopped a bit short of 21 steps. At, say, the eighteenth step I'd have invited anybody still in the game to phone in. There would have been quite a few and, with luck, one would phone. This individual would then have been invited to read out his/her written record: HTTTHHTHHHTTH HTT which would have matched the official record. So this one individual succeeded in influencing 18 successive tosses of a coin. Grasps of admiration. But admiration for what? Nothing but pure luck. I don't know if that experiment has been done. Actually, the trick here is so obvious it probably wouldn't fool many people. But how about the following?

A well-known 'psychic' goes on television, a lucrative engagement fixed up over lunch by his publicity agent. Staring out of ten million screens with hypnotically smouldering eyes (nice job by Make-up and Lighting), our imaginary seer intones that he feels a strange, spiritual rapport, a vibrating resonance of cosmic energy, with certain members of his audience. They will be able to tell who

they are because, even as he utters his mystic incantation, *their watches will stop*. After only a brief pause, a telephone on his table rings and an amplified voice in awed tones announces that its owner's watch stopped dead within seconds of the clairvoyant's words. The caller adds that she had a premonition that this was going to happen even before she looked down at her watch, for something in her hero's burning eyes seemed to speak directly to her soul. She felt the 'vibrations' of 'energy'. Even as she is speaking, a second telephone rings. Yet another watch has stopped.

A third caller's grandfather clock stopped — surely a weightier feat than stopping a little watch whose delicate hairspring would naturally be more susceptible to psychic forces than the massive pendulum of the grandfather! Another viewer's watch actually stopped a little *before* the celebrated mystic made his pronouncement — is this not an even more impressive feat of psychic control? Yet another watch has been more impatiently susceptible to occult forces. It had stopped a whole day before, *at the very moment* when its owner looked at the famous mystic's photograph in the newspaper. The studio audience gasps its appreciation. This, surely, is psychic power beyond all scepticism, for it happened a *whole day early!* 'There are more things in heaven and earth, Horatio . . .'

What we need is less gasping and more thinking. This chapter is about how to take the sting out of coincidence by quietly sitting down and calculating the likelihood that it would have happened anyway. In the course of this, we shall discover that to dismiss apparently uncanny coincidences is more interesting than gasping over them anyway.

Sometimes the calculation is easy. In a previous book I gave away the number of the combination lock on my bicycle. I felt safe in doing so because obviously my books would never be read by the kind of person who would steal a bicycle. Unfortunately somebody did steal it, and I now have a new lock with a new number, 4167. I find this number easy to remember. 41 is imprinted in my memory as the arbitrary code used to identify my clothes

and shoes at boarding school. 67 is the age at which I am due to retire. Obviously there is no interesting coincidence here: whatever the number had been, I'd have searched my life for a mnemonic recipe and I'd have found it. But mark the sequel. On the day of writing this, I received from my Oxford college a letter saying:

Each person authorized to use the photocopiers is issued with a personal code number which permits access. Your new number is 4167.

My first thought was that I'd undoubtedly lose this piece of paper (I quickly lost its equivalent last year) and I must immediately think of a formula to fix it in my memory. Something similar to the mnemonic by which I remember my bicycle combination, perhaps? So I looked again at the number on the letter and, to borrow a neat line from Fred Hoyle's science fiction novel *The Black Cloud*, the figures on the piece of paper seemed to swell to a gigantic size.

4167

I didn't need a new mnemonic. The number was identical. I rushed to tell my wife of the amazing coincidence, but on more sober reflection I shouldn't have bothered.

The odds of this happening by chance alone are easily calculated. The first digit could have been anything from 0 to 9. So there is a one in 10 chance of getting a 4 and matching the bicycle lock. For each of these ten possibilities, the second digit could have been anything from 0 to 9, so again there is a one in 10 chance of matching the bike lock's second dial. The odds of matching the first two digits is therefore one in 100 and, following the logic through the other two digits, the odds of matching all four digits of the bicycle lock is one in 10,000. It is this large number that is our protection against theft.

The coincidence is impressive. But what should we conclude? Has something mysterious and providential been going on? Have guardian angels been at work behind the scenes? Have lucky stars swum into Uranus? No. There is no reason to suspect anything more than simple accident. The number of people in the world is so large compared with 10,000 that somebody, at this very moment, is bound to be experiencing a coincidence at least as startling as mine. It just happens that today was my day to notice such a coincidence. It isn't even an added coincidence that it happened to me on this particular day, while I was writing this chapter. I had in fact written the first draft of the chapter some weeks ago. I reopened it today, after the coincidence occurred, in order to insert this anecdote. I shall surely reopen it many times to revise and polish, and I shall not remove the references to 'today': they were accurate when written. This is another way in which we habitually inflate the impressiveness of coincidence in order to make a good story.

We can do a similar calculation for the television guru whose psychic miasma seemed to stop people's watches, but we'll have to use estimates rather than exact figures. Any given watch has a certain low probability of stopping at any moment. I don't know what this probability is, but here's the kind of way in which we could come to an estimate. If we take just digital watches, their battery typically runs out within a year. Approximately, then, a digital watch stops once per year. Presumably clockwork watches stop more often because people forget to wind them and presumably digital watches stop less often because people sometimes remember to renew the battery ahead of time. But both kinds of watches probably stop as often again because they develop faults of one kind or another. So, let our estimate be that any given watch is likely to stop about once a year. It doesn't matter too much how accurate our estimate is. The principle will remain.

If somebody's watch stopped three weeks after the spell was cast, even the most credulous would prefer to put it down to chance.

We need to decide how large a delay would have been judged by the audience as sufficiently simultaneous with the psychic's announcement to impress. About five minutes is certainly safe, especially since he can keep talking to each caller for a few minutes before the next call ceases to seem roughly simultaneous. There are about 100,000 five-minute periods in a year. The probability that any given watch, say mine, will stop in a designated five-minute period is about 1 in 100,000. Low odds, but there are 10 million people watching the show. If only half of them are wearing watches, we could expect about 25 of those watches to stop in any given minute. If only a quarter of these ring in to the studio, that is 6 calls, more than enough to dumbfound a naive audience. Especially when you add in the calls from people whose watches stopped the day before, people whose watches didn't stop but whose grandfather clocks did, people who died of heart attacks and their bereaved relatives phoned in to say that their 'ticker' gave out, and so on. This kind of coincidence is celebrated in the delightfully sentimental old song, 'Grandfather's Clock':

*Ninety years without slumbering,
Tick, tock, tick, tock,
His life seconds numbering,
Tick, tock tick, tock,
It stopped . . . short . . . never to go again
When the old man died.*

Richard Feynman, in a 1963 lecture published posthumously in 1998, tells the story of how his first wife died at 9.22 in the evening and the clock in her room was later found to have stopped at exactly 9.22. There are those who would revel in the apparent mystery of this coincidence and feel that Feynman has taken away something precious when he gives us a simple, rational explanation of the mystery. The clock was old and erratic and was in the habit of stopping if tilted out of the horizontal. Feynman himself frequently

repaired it. When Mrs Feynman died, the nurse's duty was to record the exact time of death. She moved over to the clock, but it was in dark shadow. In order to see it, she picked it up – and tilted its face towards the light . . . The clock stopped. Is Feynman really spoiling something beautiful when he tells us what is surely the true – and very simple – explanation? Not for my money. For me, he is affirming the elegance and beauty of an orderly universe in which clocks stop for reasons, not to titillate human sentimental fancy.

At this point, I want to invent a technical term, and I hope you'll forgive an acronym. PETWHAC stands for Population of Events That Would Have Appeared Coincidental. Population may seem an odd word, but it is the correct statistical term. I won't keep using capital letters because they stand so unattractively on the page. Somebody's watch stopping within ten seconds of the psychic's incantation obviously belongs within the petwhac, but so do many other events. Strictly speaking, the grandfather clock's stopping should not be included. The mystic did not claim that he could stop grandfather clocks. Yet when somebody's grandfather clock did stop, they immediately telephoned in because they were, if anything, even more impressed than they would have been if their watch had stopped. The odd misconception is fostered that the psychic is even *more* powerful since he didn't even bother to mention that he could stop grandfather clocks, too! Similarly, he said nothing about watches stopping the day before or grandfathers' tickers suffering cardiac arrests.

People feel that such unanticipated events belong in the petwhac. It looks to them as though occult forces must have been at work. But when you start to think like this, the petwhac becomes really quite large, and therein lies the catch. If your watch stopped exactly 24 hours earlier, you would not have to be unduly gullible to embrace this event within the petwhac. If somebody's watch stopped exactly seven minutes before the spell, this might impress some people because seven is an ancient mystic number. And the same

presumably goes for seven hours, seven days . . . The larger the petwhac, the less we *ought* to be impressed by the coincidence when it comes. One of the devices of an effective trickster is to make people think exactly the opposite.

By the way, I deliberately chose a more impressive trick for my imaginary psychic than is actually done with watches on television. The more familiar feat is to start watches that have stopped. The television audience is invited to get up and fetch, out of drawers or attics, watches that have broken down, and hold them while the psychic performs some incantation or does some hypnotic eye work. What is really going on is that the warmth of the hand melts oil that has coagulated and this starts the watch ticking, if only briefly. Even if this happens in only a small proportion of cases, this proportion, multiplied by the large audience, will generate a satisfactory number of dumbfounded telephone calls. Actually, as Nicholas Humphrey explains in his admirable expose of supernaturalism *Soul Searching* (1995), it has been demonstrated that more than 50 per cent of broken watches start, at least momentarily, if they are held in the hand.

Here's another example of a coincidence, where it is clear how to calculate the odds. We shall use it to go on and see how odds are sensitive to changing the petwhac. I once had a girlfriend who had the same birth date (though not in the same year) as my previous girlfriend. She told a friend of hers who believed in astrology, and the friend triumphantly asked how I could possibly justify my scepticism in the face of such overwhelming evidence that I had unwittingly been brought together with two successive women on the basis of their 'stars'. Once again, let's just think it through quietly. It is easy to calculate the odds that two people, chosen entirely at random, will have the same birthday. There are 365 days in the year. Whatever the birthday of the first person, the chance that the second will have the same birthday is 1 in 365 (forgetting leap years). If we pair people off in any particular way, such as taking the successive women friends of any one man, the

odds that they will share their birthday are 1 in 365. If we take ten million men (less than the population of Tokyo or Mexico City), this apparently uncanny coincidence will have happened to more than 27,000 of them!

Now let's think about the petwhac and see how the apparent coincidence becomes less impressive as it swells. There are many other ways in which we could pair people off and still end up noticing an apparent coincidence. Two successive girlfriends with the same surname, although unrelated, for instance. Two business partners with the same birthday would also come within the petwhac; or two people with the same birthday sitting next to one another on an aeroplane. Yet, in a well-loaded Boeing 747, the odds are actually better than 50 per cent that at least one pair of neighbours will share a birthday. We don't usually notice this because we don't look over each other's shoulders as we fill in those tedious immigration forms. But if we did, somebody on most flights would go away muttering darkly about occult forces.

The birthday coincidence is famously phrased in a more dramatic way. If you have a roomful of only 23 people, mathematicians can prove that the odds are just greater than 50 per cent that at least two of them will share the same birthday. Two readers of an earlier draft of the book asked me to justify this astonishing statement. It's easier to calculate the odds that there won't be any shared birthdays and subtract from one. Forget about leap years because they're more trouble than they're worth. Suppose I bet you that with 23 people in a room, at least two will share a birthday. You bet, for the sake of argument, that there will be no shared birthdays. We're going to do the calculation by working up to 23 people gradually, starting with just one person in the room, and adding people one at a time. If at any point a match is found, I've won the bet, we stop the game and don't bother to add any more people. If we get to 23 people and there's still no match so far, you win the bet.

When the room contains only the first person, whom we may as well call A, the chance of 'No match so far' is trivially 1 (365 out of

365 chances). Now add a second person, B. The chance of a match is now one in 365. So the odds of 'No match so far' when B has joined A in the room are 364/365. Now add a third person, C. There's a one in 365 chance that C matches A and a one in 365 chance that C matches B, therefore the chance that he matches neither A nor B is 363/365 (he can't match both, because we already know that A doesn't match B). To get the total odds of 'No match so far' we have to take this 363/365 and multiply it by the odds against a match in the previous round(s), in this case by 364/365. The same reasoning applies when we add the fourth person, D. The total odds of 'No match so far' are now 364/365 \times 363/365 \times 362/365. And so on until all 23 people are in the room. Each new person adds a new term that we have to bring into the running multiplication sum, to compute 'No match so far'.

If you multiply this out to 23 terms (you have to go on down to 343/365) the answer comes out to about 0.49. This is the chance that there will not be any shared birthdays in the room. So there's a slightly greater than even chance that at least one pair of individuals in a committee of 23 will share a birthday. Most people's intuition would encourage them to bet against such a coincidence. But they'd be wrong. It is this kind of intuitive error that in general bedevils our assessment of 'uncanny' coincidences.

Here's an actual coincidence where, although it is a little harder, we can make a stab at estimating the odds approximately. My wife once bought for her mother a beautiful antique watch with a pink face. When she got it home and peeled off the price label she was amazed to find, engraved on the back of the watch, her mother's own initials, M.A.B. Uncanny? Eerie? Spine-crawling? Arthur Koestler, the famous novelist, would have read much into it. So would C. G. Jung, the widely admired psychologist and inventor of the 'collective unconscious', who also believed that a bookcase or a knife might be induced by psychic forces to explode spontaneously with a loud report. My wife, who has more sense, merely thought

the coincidence of initials remarkably convenient and sufficiently amusing to justify telling the story to me – and here I am now telling it to a wider audience.

So, what really are the odds against a coincidence of this magnitude? We can begin by calculating them in a naive way. There are 26 letters in the alphabet. If your mother has three initials and you find a watch engraved with three letters at random, the odds that the two will coincide is $1/26 \times 1/26 \times 1/26$, or one in 17,576. There are about 55 million people in Britain. If every one of them bought an antique engraved watch we'd expect more than 3,000 of them to gasp with amazement when they discovered that the watch already bore their mother's initials.

But the odds are actually better than this. Our naive calculation made the incorrect assumption that each letter has a probability of $1/26$ of being somebody's initial. This is the *average* probability for the alphabet as a whole, but some letters, such as X and Z, have a smaller probability. Others, including M, A and B, are commoner: think how much more impressed we'd be if the coinciding initials had been X.Q.Z. We can improve our estimate of odds by sampling a telephone directory. Sampling is a respectable way of estimating something that we cannot count directly. The London directory is a good place to sample because it is large and London happens to be where my wife bought the watch and where her mother lived. The London telephone directory contains about 85,000 column inches, or about 1.34 column miles, of private citizens' names. Of these, about 8,110 column inches are devoted to the Bs. This means that about 9.5 per cent of Londoners have a surname beginning with B – much more frequent than the figure for an average letter: $1/26$, or 3.8 per cent.

So, the probability that a randomly chosen Londoner would have a surname beginning with B is about 0.095 (= 9.5 per cent). What about the corresponding probabilities that the forenames will begin with M or A? It would take too long to count forename initials

right through the telephone book, and there'd be no point since the telephone book is itself only a sample. The easiest thing to do is take a subsample where forename initials are conveniently arranged in alphabetical order. This is true of the listings *within* any one surname. I shall take the commonest surname in England – Smith – and look at what proportion of the Smiths are M. Smith and what proportion are A. Smith. It is a reasonable hope that this will be approximately representative of the probabilities of forename initials for Londoners generally. It turns out that there are rather more than 20 column yards of Smiths altogether. Of these, 0.073 of them (53.6 column inches) are M. Smiths. The A. Smiths fill 75.4 column inches, representing 0.102 of all the Smiths.

If you are a Londoner and you have three initials, therefore, the chances of your initials being M.A.B. in that order are approximately $0.102 \times 0.073 \times 0.095$ or about 0.0007. Since the population of Britain is 55 million, this should mean that about 38,000 of them have the initials M.A.B., but only if everybody among those 55 million has three initials. Obviously not everybody does but, looking down the telephone directory again, it seems that at least a majority do. If we make the conservative assumption that only half of British people have three initials, that still means that more than 19,000 British people have identical initials to my wife's mother. Any one of them could have bought that watch and gasped with astonishment at the coincidence. Our calculation has shown that there is no reason to gasp.

Indeed, when we think harder about the petwhac, we find that we have even less right to be impressed. M.A.B. were the initial letters of my wife's mother's maiden name. Her married initials of M.A.W. would have seemed just as impressive had they been found on the watch. Surnames beginning with W are nearly as common in the telephone book as those beginning with B. This consideration approximately doubles the petwhac, by doubling the number of people in the country who would have been deemed, by a coincidence hunter, to have 'the same initials' as my wife's mother.

Moreover, if somebody bought a watch and found it to be engraved not with her mother's initials but with her own, she might consider it an even greater coincidence and more worthy to be embraced within the (ever-growing) petwhac.

The late Arthur Koestler, as I have already mentioned, was a great enthusiast of coincidences. Among the stories that he recounts in *The Roots of Coincidence* (1972) are several originally collected by his hero, the Austrian biologist Paul Kammerer (famous for publishing a faked experiment purportedly demonstrating the 'inheritance of acquired characteristics' in the midwife toad). Here is a typical Kammerer story quoted by Koestler:

On September 18, 1916, my wife, while waiting for her turn in the consulting rooms of Prof. Dr. J. v. H., reads the magazine Die Kunst; she is impressed by some reproductions of pictures by a painter named Schwalbach, and makes a mental note to remember his name because she would like to see the originals. At that moment the door opens and the receptionist calls out to the patients: 'Is Frau Schwalbach here? She is wanted on the telephone.'

It probably isn't worth trying to estimate the odds against this coincidence, but we can at least write down some of the things that we'd need to know. 'At that moment the door opens' is a little vague. Did the door open one second after she made the mental note to look up Schwalbach's paintings or 20 minutes? How long could the interval have been, leaving her still impressed by the coincidence? The frequency of the name Schwalbach is obviously relevant: we'd be less impressed if it had been Schmidt or Strauss; more impressed if it had been Twistleton-Wykeham-Fiennes or Knatchbull-Hugueson. My local library doesn't have the Vienna telephone book, but a quick look in another large Germanic telephone directory, the Berlin one, yields half a dozen Schwalbachs: the name is not particularly common, therefore, and it is understandable that the lady was impressed. But we need to think further about

the size of the petwhac. Similar coincidences could have happened to people in other doctors' waiting rooms; and in dentists' waiting rooms, government offices and so on; and not just in Vienna but anywhere else. The quantity to keep bearing in mind is the number of *opportunities* for coincidence that would have been thought, if they had occurred, just as remarkable as the one that actually did occur.

Now let's take another kind of coincidence, where it is even harder to know how to start calculating odds. Consider the often-quoted experience of dreaming of an old acquaintance for the first time in years and then getting a letter from him, out of the blue, the next day. Or of learning that he died in the night. Or of learning that he didn't die in the night but his father did. Or that his father didn't die but won the football pools. See how the petwhac grows out of control when we relax our vigilance?

Often, these coincidence stories are gathered together from a large field. The correspondence columns of popular newspapers contain letters sent in by individual readers who would not have written but for the amazing coincidence that had happened to them. In order to decide whether we should be impressed, we need to know the circulation figure for the newspaper. If it is 4 million, it would be surprising if we did *not* read daily of some stunning coincidence, since a coincidence only has to happen to one of the 4 million in order for us to have a good chance of being told about it in the paper. It is hard to calculate the probability of a particular coincidence happening to one person, say a long-forgotten old friend dying during the night we happen to dream about him. But whatever this probability is, it is surely far greater than one in 4 million.

So, there really is no reason for us to be impressed when we read in the newspaper of a coincidence that has happened to one of the readers, or to somebody, somewhere in the world. This argument against being impressed is entirely valid. Nevertheless, there may be something lurking here that still bothers us. You may be happy to agree that, from the point of view of a reader of a mass-circulation

newspaper, we have no right to be impressed at a coincidence that happens to another of the millions of readers of the same newspaper who bothers to write in. But it is much harder to shake the feeling of spine-chilled awe when the coincidence happens to *you yourself*. This is not just personal bias. One can make a serious case for it. The feeling occurs to almost everybody I meet; if you ask anybody at random, there is a good chance that they will have at least one pretty uncanny story of coincidence to relate. On the face of it, this undermines the sceptic's point about newspaper stories having been culled from a millions-strong readership – a huge catchment of opportunity.

Actually it doesn't undermine it, for the following reason. Each one of us, though only a single person, none the less amounts to a very large population of *opportunities* for coincidence. Each ordinary day that you or I live through is an unbroken sequence of events, or incidents, any of which is potentially a coincidence. I am now looking at a picture on my wall of a deep-sea fish with a fascinatingly alien face. It is possible that, at this very moment, the telephone will ring and the caller will identify himself as a Mr Fish. I'm waiting . . .

The telephone didn't ring. My point is that, whatever you may be doing in any given minute of the day, there probably is some other event – a phone call, say – which, if it were to happen, would with hindsight be rated an eerie coincidence. There are so many minutes in every individual's lifetime that it would be quite surprising to find an individual who had *never* experienced a startling coincidence. During this particular minute, my thoughts have strayed to a schoolfellow called Haviland (I don't remember his Christian name, nor what he looked like) whom I haven't seen or thought of for 45 years. If, at this moment, an aeroplane manufactured by the de Havilland company were to fly past the window, I'd have a coincidence on my hands. In fact I have to report that no such plane has been forthcoming, but I have now moved on to think about something else, which gives yet another opportunity

for coincidence. And so the opportunities for coincidence go on throughout the day and every day. But the negative occurrences, the failures to coincide, are not noticed and not reported.

Our propensity to see significance and pattern in coincidence, whether or not there is any real significance there, is part of a more general tendency to seek patterns. This tendency is laudable and useful. Many events and features in the world really are patterned in a non-random way and it is helpful to us, and to animals generally, to detect these patterns. The difficulty is to navigate between the Sirens of detecting apparent pattern when there isn't any, and the Charybdis of failing to detect pattern when there is.

The science of statistics is quite largely concerned with steering this difficult course. But long before statistical methods were formalized, humans and indeed other animals were reasonably good intuitive statisticians. It is easy to make mistakes, however, in both directions.

Here are some true statistical patterns in nature which are not totally obvious, and which humans have not always known.

True pattern

Sexual intercourse is statistically followed by birth about 266 days later

Reason difficult to detect
The exact interval varies around the average of 266 days. Intercourse more often than not fails to result in conception. Intercourse is often frequent anyway, so it is not obvious that conception results from that rather than from, say, eating, which is also frequent.

Conception is relatively probable in the middle of a woman's cycle, and relatively improbable menstruation

UNWEAVING THE UNCANNY	UNWEAVING THE RAINBOW
Smoking causes lung cancer	Plenty of people who smoke don't get lung cancer. Many people get lung cancer who never smoked.
In a time of bubonic plague, proximity to rats, and especially their fleas, tends to lead to infection	Lots of rats and fleas around anyway. Rats and fleas are associated with so many other things, such as dirt and 'bad air', and it is hard to know which of the many correlated factors is the important one. I.e. again, there are spurious correlations that get in the way.
Droughts can be brought to an end by a rain dance (or human sacrifice, or sprinkling goats' blood on a ferret's kidneys, or whatever arbitrary custom the particular theology lays down)	<i>False pattern</i> Occasionally, rains do chance to follow upon a rain dance (etc.), and these rare lucky strikes lodge in the memory. When the rain dance, say, is not followed by rain, it is assumed that some detail went wrong with the ceremony, or that the gods are angry for some other reason: it is always easy enough to find a sufficiently plausible excuse.
See above. In addition, women who don't menstruate don't conceive. This is a spurious correlation which gets in the way and even, to a naive mind, suggests the opposite of the truth.	<i>Reason easy to be misled</i> See above. Also, it is in the interests of astrologers to foster the myth, just as it is no doubt in the interests of priests and witch-doctors to foster the myths about rain dances and ferrets' kidneys.

After a run of ill-luck, good luck becomes more likely

If bad luck persists, we assume that the run of bad luck hasn't ended yet, and we look forward all the more to its eventual end. If bad luck does not persist, the prophecy is seen as fulfilled. We subconsciously *define* a 'run' of bad luck in terms of its end. Therefore it obviously has to be followed by good luck.

We are not the only animals to seek statistical patterns of non-randomness in nature, and we are not the only animals to make mistakes of the kind that might be called superstitious. Both these facts are neatly demonstrated in the apparatus called the Skinner box, after the famous American psychologist B. F. Skinner. A Skinner box is a simple but versatile piece of equipment for studying the psychology of, usually, a rat or a pigeon. It is a box with a switch or switches let into one wall which the pigeon (say) can operate by pecking. There is also an electrically operated feeding (or other rewarding) apparatus. The two are connected in such a way that pecking by the pigeon has some influence on the feeding apparatus. In the simplest case, every time the pigeon pecks the key it gets food. Pigeons readily learn the task. So do rats and, in suitably enlarged and reinforced Skinner boxes, so do pigs.

We know that the causal link between key peck and food is provided by electrical apparatus, but the pigeon doesn't. As far as the pigeon is concerned, pecking a key might as well be a rain dance. Moreover, the link can be quite a weak, statistical one. The apparatus may be set up so that, far from every peck being rewarded, only one in 10 pecks is rewarded. This can mean literally every tenth peck. Or, with a different setting of the apparatus, it can mean that one in 10 pecks *on average* is rewarded, but on any particular occasion the exact number of pecks required is determined at random. Or there may be a clock which determines that one tenth of the time, on average, a peck will yield reward, but it is

impossible to tell which tenth of the time. Pigeons and rats learn to press keys even when, one might think, you'd need to be quite a good statistician to detect the cause-effect relationship. They can be worked up to a schedule in which only a very small proportion of pecks is rewarded. Interestingly, habits learned when pecks are only occasionally rewarded are more durable than habits learned when all pecks are rewarded: the pigeon is less swiftly discouraged when the rewarding mechanism is switched off altogether. This makes intuitive sense if you think about it.

Pigeons and rats, then, are quite good statisticians, able to pick up slight, statistical laws of patterning in their world. Presumably this ability serves them in nature as well as in the Skinner box. Life out there is rich in pattern, the world is a big, complicated Skinner box. Actions by a wild animal are frequently followed by rewards or punishments or other important events. The relationship between cause and effect is frequently not absolute but statistical. If a curlew probes mud with its long, curved bill, there is a certain probability that it will strike a worm. The relationship between probe events and worm events is statistical but real. A whole school of research on animals has grown up around so-called Optimal Foraging Theory. Wild birds show quite sophisticated abilities to assess, statistically, the relative food-richness of different areas and they switch their time between the areas accordingly.

Back in the laboratory, Skinner founded a large school of research using Skinner boxes for all kinds of detailed purposes. Then, in 1948, he tried a brilliant variant on the standard technique. He completely severed the causal link between behaviour and reward. He set up the apparatus to 'reward' the pigeon from time to time *no matter what the bird did*. Now all that the birds actually needed to do was sit back and wait for the reward. But in fact this is not what they did. Instead, in six out of eight cases, they built up – exactly as though they were learning a rewarded habit – what Skinner called 'superstitious' behaviour. Precisely what this consisted of varied from pigeon to pigeon. One bird spun itself round

like a top, two or three turns anticlockwise, between 'rewards'. Another bird repeatedly thrust its head towards one particular upper corner of the box. A third bird showed 'tossing' behaviour, as if lifting an invisible curtain with its head. Two birds independently developed the habit of rhythmic, side-to-side 'pendulum swinging' of the head and body. This last habit, incidentally, must have looked rather like the courtship dance of some birds of paradise. Skinner used the word superstition because the birds behaved as if they thought that their habitual movement had a causal influence on the reward mechanism, when actually it didn't. It was the pigeon equivalent of a rain dance.

A superstitious habit, once established, might persist for hours, long after the reward mechanism had been switched off. The habits did not, however, remain unchanged in form. They drifted, like the progressive improvisations of an organist. In one typical case the pigeon's superstitious habit began as a sharp movement of the head from the middle position towards the left. As time went by, the movement became more energetic. Eventually the whole body moved in the same direction and a step or two would be taken with the legs. After many hours of 'topographic drift', this leftward stepping movement became the predominant feature of the habit. The superstitious habits themselves may have been derived from the species' natural repertoire, but it is still fair to say that performing them in this context, and performing them repeatedly, is unnatural for pigeons.

Skinner's superstitious pigeons were behaving like statisticians, but statisticians who have got it wrong. They were alert to the possibility of links between events in their world, especially links between rewards that they wanted and actions that it was in their power to take. A habit, such as shoving the head up into the corner of the cage, began by chance. The bird just happened to do it at the moment before the reward mechanism was due to clunk into action. Understandably enough, the bird developed the tentative hypothesis that there was a link between the two events. So it

shoved its head into the corner again. Sure enough, by the luck of Skinner's timing mechanism, the reward came again. If the bird had tried the experiment of *not* shoving its head into the corner, it would have found that the reward came anyway. But it would have needed to be a better and more sceptical statistician than many of us humans are in order to try this experiment.

Skinner makes the comparison with human gamblers developing little lucky 'tics' when playing cards. This kind of behaviour is also a familiar spectacle on bowling greens. Once the 'wood' (ball) has left the bowler's hand there is nothing more he can do to encourage it to move towards the 'jack' (target ball). Nevertheless, expert bowlers nearly always trot after their wood, often still in the stooped position, twisting and turning their bodies as if to impart desperate instructions to the now indifferent ball, and often speaking futile words of encouragement to it. A one-arm bandit in Las Vegas is nothing more nor less than a human Skinner box. 'Key-pecking' is represented not just by pulling the lever but also, of course, by putting money in the slot. It really is a fool's game because the odds are known to be stacked in favour of the casino – how else would the casino pay its huge electricity bills? Whether or not a given lever pull will deliver a jackpot is determined at random. It is a perfect recipe for superstitious habits. Sure enough, if you watch gambling addicts in Las Vegas you see movements highly reminiscent of Skinner's superstitious pigeons. Some talk to the machine. Others make funny signs to it with their fingers, or stroke it or pat it with their hands. They once patted it and won the jackpot and they've never forgotten it. I have watched computer addicts, impatient for a server to respond, behaving in a similar way, say, knocking the terminal with their knuckles.

My informant about Las Vegas has also made an informal study of London betting shops. She reports that one particular gambler habitually runs, after placing his bet, to a certain tile in the floor, where he stands on one leg while watching the race on the bookmaker's television. Presumably he once won while standing

on this tile and conceived the notion that there was a causal link. Now, if somebody else stands on 'his' lucky tile (some other sportsmen do this deliberately, perhaps to try to hijack some of his 'luck' or just to annoy him) he dances around it, desperately trying to get a foot on the tile before the race ends. Other gamblers refuse to change their shirt, or to cut their hair, while they are 'on a lucky streak'. In contrast one Irish punter, who had a fine head of hair, shaved it completely bald in a desperate effort to change his luck. His hypothesis was that he was having rotten luck on the horses and he had lots of hair. Perhaps the two were connected somehow; perhaps these facts were all part of a meaningful pattern! Before we feel too superior, let us remember that large numbers of us were brought up to believe that Samson's fortunes changed utterly after Delilah cut off his hair.

How can we tell which apparent patterns are genuine, which random and meaningless? Methods exist, and they belong in the science of statistics and experimental design. I want to spend a little more time explaining a few of the principles, though not the details, of statistics. Statistics can largely be seen as the art of distinguishing pattern from randomness. Randomness means lack of pattern. There are various ways of explaining the ideas of randomness and pattern. Suppose I claim that I can tell girls' handwriting from boys'. If I am right, this would have to mean that there is a real pattern relating sex to handwriting. A sceptic might doubt this, agreeing that handwriting varies from person to person but denying that there is a sex-related pattern to this variation. How should you decide whether my claim, or the sceptic's, is right? It is no use just accepting my word for it. Like a superstitious Las Vegas gambler, I could easily have mistaken a lucky streak for a real, repeatable skill. In any case, you have every right to demand evidence. What evidence should satisfy you? The answer is evidence that is publicly recorded, and properly analysed.

The claim is, in any case, only a statistical claim. I do not maintain (in this hypothetical example – in reality I am not

claiming anything) that I can infallibly judge the sex of the author of a given piece of handwriting. I claim only that among the great variation that exists among handwriting, some component of that variation correlates with sex. Therefore, even though I shall often make mistakes, if you give me, say, 100 samples of handwriting I should be able to sort them into boys and girls more accurately than could be achieved purely by guessing at random. It follows that, in order to assess my claim, you are going to have to calculate how likely it is that a given result could have been achieved by guessing at random. Once again, we have an exercise in calculating the odds of coincidence.

Before we get to the statistics, there are some precautions you need to take in designing the experiment. The pattern – the non-randomness we seek – is a pattern relating sex to handwriting. It is important not to confound the issue with extraneous variables. The handwriting samples that you give me should not, for instance, be personal letters. It would be too easy for me to guess the sex of the writer from the content of the letter rather than from the handwriting. Don't choose all the girls from one school and all the boys from another. The pupils from one school might share aspects of their handwriting, learning either from each other or from a teacher. These could result in real differences in handwriting, and they might even be interesting, but they could be representative of different schools, and only incidentally of different sexes. And don't ask the children to write out a passage from a favourite book. I should be influenced by a choice of *Black Beauty* or *Biggles* (readers whose childhood culture is different from mine will substitute examples of their own).

Obviously, it is important that the children should all be strangers to me, otherwise I'd recognize their individual writing and hence know their sex. When you hand me the papers they must not have the children's names on them, but you must have some means of keeping track of whose is which. Put secret codes on them for your own benefit, but be careful how you choose the codes. Don't put a

green mark on the boys' papers and a yellow mark on the girls'. Admittedly, I won't know which is which, but I'll guess that yellow denotes one sex and green the other, and that would be a big help. It would be a good idea to give every paper a code number. But don't give the boys the numbers 1 to 10 and the girls 11 to 20; that would be just like the yellow and green marks all over again. So give the papers *random* numbers and keep the crib list locked up where I cannot find it. These precautions are those named 'double blind' in the literature of medical trials.

Let's assume that all the proper double blind precautions have been taken, and that you have assembled 20 anonymous samples of handwriting, shuffled into random order. I go through the papers, sorting them into two piles for suspected boys and suspected girls. I may have some 'don't knows', but let's assume that you compel me to make the best guess I can in such cases. At the end of the experiment I have made two piles and you look through to see how accurate I have been.

Now the statistics. You'd expect me to guess right quite often even if I was guessing purely at random. But *how* often? If my claim to be able to sex handwriting is unjustified, my guessing rate should be no better than somebody tossing a coin. The question is whether my actual performance is sufficiently different from a coin-tosser's to be impressive. Here is how to set about answering the question.

We have to set 211 against the total number of ways in which the 20 scripts could have been classified by penny-tossers. This is not difficult to calculate. The first script could have been boy or girl: that is two possibilities. The second script also could have been boy or girl. So, for each of the two possibilities for the first script, there were two possibilities for the second. That is $2 \times 2 = 4$ possibilities for the first two scripts. The possibilities for the first three scripts are $2 \times 2 \times 2 = 8$. And the possible ways of classifying all 20 scripts are $2 \times 2 \times 2 \dots 20$ times, or 2 to the power 20. This is a pretty big number, 1,048,576.

Think about all *possible* ways in which I could have guessed the sex of the 20 writers. List them in order of impressiveness, beginning with all 20 correct and going down to completely random (all 20 exactly wrong is nearly as impressive as all 20 exactly right, because it shows that I can discriminate, even though I perversely reverse the sign). Then look at the *actual* way I sorted them and count up the percentage of all possible sortings that would have been as impressive as the actual one, or more. Here's how to think about all possible sortings. First, note that there is only one way of being

So, of all possible ways of guessing, the proportion of ways that are as good as or better than my actual score is 211 divided by 1,048,576, which is approximately 0.0002, or 0.02 per cent. To put it another way, if 10,000 people sorted the scripts entirely by tossing pennies, you'd expect only two of them to score as well as I actually did. This means that my score is pretty impressive and, if I performed as well as this, it would be strong evidence that boys and girls differ systematically in their handwriting. Let me repeat that this is all hypothetical. As far as I know, I have no such ability to sex handwriting. I should also add that, even if there was good evidence for a sex difference in handwriting, this would say nothing about whether the difference is innate or learned. The evidence, at least if it came from the kind of experiment just described, would be equally compatible with the idea that girls are systematically taught a different handwriting from boys — perhaps a more 'ladylike' and less 'assertive' fist.

We have just performed what is technically called a test of statistical significance. We reasoned from first principles, which made it rather tedious. In practice, research workers can call upon tables of probabilities and distributions that have been previously calculated. We therefore don't literally have to write down all possible ways in which things could have happened. But the underlying theory, the basis upon which the tables were calculated, depends, in essence, upon the same fundamental procedure. Take the events that *could* have been obtained and throw them down repeatedly at random. Look at the *actual* way the events occurred and measure how extreme it is, on the spectrum of all possible ways in which they could have been thrown down.

Notice that a test of statistical significance does not prove anything conclusively. It can't rule out luck as the generator of the result that we observe. The best it can do is place the observed result on a par with a specified amount of luck. In our particular hypothetical example, it was on a par with two out of 10,000 random guessers. When we say that an effect is statistically significant, we must

always specify a so-called p-value. This is the probability that a purely random process would have generated a result at least as impressive as the actual result. A p-value of 2 in 10,000 is pretty impressive, but it is still possible that there is no genuine pattern there. The beauty of doing a proper statistical test is that we know how probable it is that there is no genuine pattern there.

Conventionally, scientists allow themselves to be swayed by p-values of 1 in 100, or even as high as 1 in 20: far less impressive than 2 in 10,000. What p-value you accept depends upon how important the result is, and upon what decisions might follow from it. If all you are trying to decide is whether it is worth repeating the experiment with a larger sample, a p-value of 0.05, or 1 in 20, is quite acceptable. Even though there is a 1 in 20 chance that your interesting result would have happened anyway by chance, not much is at stake: the error is not a costly one. If the decision is a life and death matter, as in some medical research, a much lower p-value than 1 in 20 should be sought. The same is true of experiments that purport to show highly controversial results, such as telepathy or 'paranormal' effects.

As we briefly saw in connection with DNA fingerprinting, statisticians distinguish false positive from false negative errors, sometimes called type 1 and type 2 errors respectively. A type 2 error, or false negative, is a failure to detect an effect when there really is one. A type 1 error, or false positive, is the opposite: concluding that there really is something going on when actually there is nothing but randomness. The p-value is the measure of the probability that you have made a type 1 error. Statistical judgement means steering a middle course between the two kinds of error. There is a type 3 error in which your mind goes totally blank whenever you try to remember which is which of type 1 and type 2. I still look them up after a lifetime of use. Where it matters, therefore, I shall use the more easily remembered names, false positive and false negative. I also, by the way, frequently make mistakes in arithmetic. In practice I should never dream of doing

a statistical test from first principles as I did for the hypothetical handwriting case. I'd always look up in a table that somebody else – preferably a computer – had calculated.

Skinner's superstitious pigeons made false positive errors. There was in fact no pattern in their world that truly connected their actions to the deliveries of the reward mechanism. But they behaved as if they had detected such a pattern. One pigeon 'thought' (or behaved as if it thought) that left stepping caused the reward mechanism to deliver. Another 'thought' that thrusting its head into the corner had the same beneficial effect. Both were making false positive errors. A false negative error is made by a pigeon in a Skinner box who never notices that a peck at the key yields food if the red light is on, but that a peck when the blue light is on punishes by switching the mechanism off for ten minutes. There is a genuine pattern waiting to be detected in the little world of this Skinner box, but our hypothetical pigeon does not detect it. It pecks indiscriminately to both colours, and therefore gets a reward less frequently than it could.

A false positive error is made by a farmer who thinks that sacrificing to the gods brings longed-for rain. In fact, I presume (although I haven't investigated the matter experimentally), there is no such pattern in his world, but he does not discover this and persists in his useless and wasteful sacrifices. A false negative error is made by a farmer who fails to notice that there is a pattern in the world relating manuring of a field to the subsequent crop yield of that field. Good farmers steer a middle way between type 1 and type 2 errors.

It is my thesis that all animals, to a greater or lesser extent, behave as intuitive statisticians, choosing a middle course between type 1 and type 2 errors. Natural selection penalizes both type 1 and type 2 errors, but the penalties are not symmetrical and no doubt vary with the different ways of life of species. A stick caterpillar looks so like the twig it is sitting on that we cannot doubt that natural selection has shaped it to resemble a twig. Many

caterpillars died to produce this beautiful result. They died because they did not sufficiently resemble a twig. Birds, or other predators, found them out. Even some very good twig mimics must have been found out. How else did natural selection push evolution towards the pitch of perfection that we see? But, equally, birds must many times have missed caterpillars because they resembled twigs, in some cases only slightly. Any prey animal, no matter how poorly camouflaged, can be detected by predators under ideal seeing conditions. Equally, any prey animal, no matter how poorly camouflaged, can be missed by predators under bad seeing conditions. Seeing conditions can vary with angle (a predator may spot a well-camouflaged animal when looking straight at it, but will miss a poorly camouflaged animal out of the corner of its eye). They can vary with light intensity (a prey may be overlooked at twilight, whereas it would be seen at noon). They can vary with distance (a prey which would be seen at six inches range may be overlooked at a range of 100 yards).

Imagine a bird cruising around a wood, looking for prey. It is surrounded by twigs, a very few of which might be edible caterpillars. The problem is to decide. We can assume that the bird could guarantee to tell whether an apparent twig was actually a caterpillar if it approached the twig really close and subjected it to a minute, concentrated examination in a good light. But there isn't time to do that for all twigs. Small birds with high turnover metabolism have to find food alarmingly often in order to stay alive. Any bird that scanned every individual twig with the equivalent of a magnifying glass would die of starvation before it found its first caterpillar. Efficient searching demands a faster, more cursory and rapid scanning, even though this carries a risk of missing some food. The bird has to strike a balance. Too cursory and it will never find anything. Too detailed and it will detect every caterpillar it looks at, but it will look at too few, and starve.

It is easy to apply the language of type 1 and type 2 errors. A false negative is committed by a bird that sails by a caterpillar

without giving it a closer look. A false positive is committed by a bird that zooms in on a suspected caterpillar, only to discover that it is really a twig. The penalty for a false positive is the time and energy wasted flying in for the close inspection: not serious on any one occasion, but it could mount up fatally. The penalty for a false negative is missing a meal. No bird outside Cloud Cuckooland can hope to be free of all type 1 and type 2 errors. Individual birds will be programmed by natural selection to adopt some compromise policy calculated to achieve an optimum intermediate level of false positives and false negatives. Some birds may be biased towards type 1 errors, others towards the opposite extreme. There will be some intermediate setting which is best, and natural selection will steer evolution towards it.

Which intermediate setting is best will vary from species to species. In our example it will also depend upon conditions in the wood, for example, the size of the caterpillar population in relation to the number of twigs. These conditions may change from week to week. Or they may vary from wood to wood. Birds may be programmed to learn to adjust their policy as a result of their statistical experience. Whether they learn or not, successfully hunting animals must usually behave as if they are good statisticians. (I hope it is not necessary, by the way, to plod through the usual disclaimer: No, no, the birds aren't consciously working it out with calculator and probability tables. They are behaving *as if* they were calculating p-values. They are no more aware of what a p-value means than you are aware of the equation for a parabolic trajectory when you catch a cricket ball or baseball in the outfield.)

Angler fish take advantage of the gullibility of little fish such as gobies. But that is an unfairly value-laden way of putting it. It would be better not to speak of gullibility and say that they exploit the inevitable difficulty the little fish have in steering between type 1 and type 2 errors. The little fish themselves need to eat. What they eat varies, but it often includes small wriggling objects such as worms or shrimps. Their eyes and nervous systems are

tuned to wriggling things. They look for wriggling movement and if they see it they pounce. The angler fish exploits this tendency. It has a long fishing rod, evolved from a modified spine, commanded by natural selection from its original location at the front of the dorsal fin. The angler fish itself is highly camouflaged and sits motionless on the sea bottom for hours at a time, blending perfectly with the weeds and rocks. The only part of it which is conspicuous is a 'bait', which looks like a worm, a shrimp or a small fish, at the end of its fishing rod. In some deep-sea species the bait is even luminous. In any case, it seems to wriggle like something worth eating when the angler waves its rod. A possible prey fish say, a goby, is attracted. The angler 'plays' its prey for a little while to hook its attention, then casts the bait down into the still unsuspected region in front of its own invisible mouth, and the little fish often follows. Suddenly that huge mouth is invisible no longer. It gapes massively, there is a violent inrushing of water, engulfing every floating object in the vicinity, and the little fish has pursued its last worm.

From the point of view of a hunting goby, any worm may be overlooked or it may be seen. Once the 'worm' has been detected, it may turn out to be a real worm or an angler fish's lure, and the unfortunate fish is faced with a dilemma. A false negative error would be to refrain from attacking a perfectly good worm for fear that it might be an angler fish lure. A false positive error would be to attack a worm, only to discover that it is really a lure. Once again, it is impracticable in the real world to get it right all the time. A fish that is too risk-averse will starve because it never attacks worms. A fish that is too foolhardy won't starve but it may be eaten. The optimum in this case may not be halfway between. More surprisingly, the optimum may be one of the extremes. It is possible that angler fish are sufficiently rare that natural selection favours the extreme policy of attacking all apparent worms. I am fond of a remark of the philosopher and psychologist William James on human angling:

There are more worms unattached to hooks than impaled upon them; therefore, on the whole, says Nature to her fishy children, bite at every worm and take your chances.

(1910)

Like all other animals, and even plants, humans can and must behave as intuitive statisticians. The difference with us is that we can do our calculations twice over. The first time intuitively, as though we were birds or fish. And then again explicitly, with pencil and paper or computer. It is tempting to say that the pencil and paper way gets the right answer, so long as we don't make some publicly detectable blunder like adding in the date, whereas the intuitive way may yield the wrong answer. But there strictly is no 'right' answer, even in the case of pencil and paper statistics. There may be a right way to do the sums, to calculate the p-value, but the criterion, or threshold p-value, that we demand before choosing a particular action is still our decision and it depends upon our aversion to risk. If the penalty for making a false positive error is much greater than the penalty for making a false negative error, we should adopt a cautious, conservative threshold: almost never try a 'worm' for fear of the consequences. Conversely, if the risk-symmetry is opposite, we should rush in and try every 'worm' that is going: it is unlikely to matter if we keep tasting false worms so we may as well have a go.

Taking on board the need to steer between false positive and false negative errors, let me return to uncanny coincidence and the calculation of the probability that it would have happened anyway. If I dream of a long-forgotten friend who dies the same night, I am tempted, like anyone else, to see meaning or pattern in the coincidence. I really have to force myself to remember that quite a few people die every night, masses of people dream every night, they quite often dream that people die, and coincidences like this are probably happening to several hundred people in the world every night. Even as I think this through, my own intuition cries out that there must be meaning in the coincidence because it has

happened to me. If it is true that intuition is, in this case, making a false positive error, we need to come up with a satisfactory explanation for why human intuition errs in this direction. As Darwinians, we should be alive to the possible pressures towards erring on the type 1 or the type 2 side of the divide.

As a Darwinian, I want to suggest that our willingness to be impressed at apparently uncanny coincidence (which is a case of our willingness to see pattern where there is none) is related to the typical population size of our ancestors and the relative poverty of their everyday experience. Anthropology, fossil evidence and the study of other apes all suggest that our ancestors, for much of the past few million years, probably lived in either small roving bands or small villages. Either of these would mean that the number of friends and acquaintances that our ancestors would ordinarily meet and talk to with any frequency was not more than a few dozen. A prehistoric villager could expect to hear stories of startling coincidence in proportion to this small number of acquaintances. If the coincidence happened to somebody not in his village, he wouldn't hear the story. So our brains became calibrated to detect pattern and grasp with astonishment at a level of coincidence which would actually be quite modest if our catchment area of friends and acquaintances had been large.

Nowadays, our catchment area is large, especially because of newspapers, radio and other vehicles of mass news circulation. I've already spelled out the argument. The very best and most spine-creeping coincidences have the opportunity to circulate, in the form of bated-breath stories, over a far wider audience than was ever possible in ancestral times. But, I am now conjecturing, our brains are calibrated by ancestral natural selection to expect a much more modest level of coincidence, calibrated under small village conditions. So we are impressed by coincidences because of a miscalibrated grasp threshold. Our subjective petwhacs have been calibrated by natural selection in small villages, and, as is the case with so much of modern life, the calibration is now out of date. (A

similar argument could be used to explain why we are so hysterically risk-averse to hazards that are much publicized in the newspapers—perhaps anxious parents who imagine ravening paedophiles lurking behind every lamp post on their children's walk from school are 'miscalibrated'.

I guess that there may be another, particular effect pushing in the same direction. I suspect that our individual lives under modern conditions are richer in experiences per hour than were ancestral lives. We don't just get up in the morning, scratch a living in the same way as yesterday, eat a meal or two and go to sleep again. We read books and magazines, we watch television, we travel at high speed to new places, we pass thousands of people in the street as we walk to work. The number of faces we see, the number of different situations we are exposed to, the number of separate things that happen to us, is much greater than for our village ancestors. This means that the number of opportunities for coincidence is greater for each one of us than it would have been for our ancestors, and consequently greater than our brains are calibrated to assess. This is an additional effect, over and above the population size effect that I have already noted.

With respect to both these effects, it is theoretically possible for us to recalibrate ourselves, learn to adjust our *gasp* threshold to a level more appropriate to modern populations and modern richnesses of experience. But this seems to be revealingly difficult even for sophisticated scientists and mathematicians. The fact that we still do *gasp* when we do, that clairvoyants and mediums and psychics and astrologers manage to make such a nice living out of us, all suggests that we do not, on the whole, learn to recalibrate ourselves. It suggests that the parts of our brains responsible for doing intuitive statistics are still back in the stone age.

The same may be true of intuition generally. In *The Unnatural Nature of Science* (1992), the distinguished embryologist Lewis Wolpert has argued that science is difficult because it is more or less systematically counter-intuitive. This is contrary to the view

of T. H. Huxley (Darwin's Bulldog) who saw science as 'nothing but trained and organized common sense, differing from the latter only as a veteran may differ from a raw recruit'. For Huxley, the methods of science 'differ from those of common sense only as far as the guardsman's cut and thrust differ from the manner in which a savage wields his club'. Wolpert insists that science is deeply paradoxical and surprising; an affront to common sense rather than an extension of it, and he makes a good case. For example, every time you drink a glass of water you are imbibing at least one molecule that passed through the bladder of Oliver Cromwell. This follows by extrapolation from Wolpert's observation that 'there are many more molecules in a glass of water than there are glasses of water in the sea'. Newton's law that objects stay in motion unless positively stopped is counter-intuitive. So is Galileo's discovery that, when there is no air resistance, light objects fall at the same rate as heavy objects. So is the fact that solid matter, even a hard diamond, consists almost entirely of empty space. Steven Pinker gives an illuminating discussion of the evolutionary origins of our physical intuitions in *How the Mind Works* (1998).

More profoundly difficult are the conclusions of quantum theory, overwhelmingly supported by experimental evidence to a stupefyingly convincing number of decimal places, yet so alien to the evolved human mind that even professional physicists don't understand them in their intuitive thoughts. It seems to be not just our intuitive statistics but our very minds themselves that are back in the stone age.