

Probability Paradoxes: an improbable journey to the End of the World

Rose D Baker C. Math FIMA

Centre for Operational Research and Applied Statistics, University of Salford

Probability theory abounds with startling and paradoxical results. They are described here under three headings, ranging from puzzles and brainteasers through to the philosophically perplexing. After a brief tour of some familiar and less-familiar paradoxes, the reader is confronted with the Doomsday paradox, and subsequently subjected to a sadistic torture in the 'Devil's shooting room'.

Puzzles

Probability theory is a dry subject to most students, although its study was begun by some very colourful individuals such as Cardano [9], the 16th century gambling scholar, and was motivated by gaming problems far removed from academia. Probability paradoxes and startling results in fact abound, which I divide threefold into puzzles, intriguing results, and the philosophically deep.

A paradox is a result that conflicts with common belief. The simplest probability paradoxes make good puzzles and brainteasers, and have been well publicised. One that has become popular in recent years is the Monty Hall paradox or dilemma. Monty Hall told contestants in his game show that a car was concealed behind one of three doors. Contestants had to guess which door. Monty then opened a door *not* containing the car, and that contestants had not chosen, and they were allowed to switch guesses between the two remaining doors. Should they do so or not? Most people intuitively think not, but the paradox is that they should, and they will then raise their chance of success from $1/3$ to $2/3$.

Several others are close relations to this one, such as the three prisoners' problem or dilemma. Here one of three prisoners

is to be executed, and the guard offers to tell prisoner A the name of another of the three who is *not* to be executed. Our man is puzzled: if he accepts the offer, has he increased his probability of execution from $1/3$ to $1/2$? Note that analogously to the Monty Hall problem, what has actually happened is that the *other* prisoner's probability has increased to $2/3$. Continuing the theme of three, in what has been called the 'three pancakes problem', we have three cards, one green on both sides, one red on both sides, and one red on one side and green on the other. A card is randomly selected and thrown onto the floor. If the side uppermost is red, what is the probability that this is the red card? The mystery here is of course where the pancakes come in.

In one version of the related Lewis Carroll's pillow problem [3], a black or white die is put randomly into a bag. A white die is added, the bag is shaken, and a die removed without looking, which, it turns out, is white. What is the probability that the remaining die is white? What turns this one from a simple probability problem into a paradox is that the state of the bag is unchanged, so we may feel that the answer should be $1/2$ (it is not). This paradox thus complements the three prisoners problem.

I won't waste space or spoil all the fun by giving all the answers, but looking up the words 'probability' and 'paradox' with Google will give websites with information on these puzzles and many more. The Monty Hall problem, and several other paradoxes, are given in Barbeau (2000). I can not resist giving one more paradox quoted by him, known as the two envelopes paradox or the box problem. Two envelopes contain money, one twice as much as the other. One envelope is assigned to you by a coin toss. One can

imagine that the envelope is opened, and contains say £10, or that it remains unopened. Should you switch envelopes? If your envelope contains an amount £A, the other envelope is presumably equally likely to contain £2A or £A/2, and your average gain on switching would be £A/4, so perhaps you should switch? Or not.

You are equally likely to have the higher or the lower amount of money, that situation will not change on switching envelopes, and so you need not switch (this conclusion would change if further information were available, for example if a ceiling on the amounts in the envelopes were known). What should be worrying to statisticians is that a seemingly obvious Bayesian argument gives wrong results here, leading presumably to an endless switching of envelopes. The simple Bayesian argument about the content of the other envelope in fact applies to a different situation, one in which you are always given the same sum of money, and the other envelope, on the toss of a coin, contains either half or twice as much. Here you should switch, unless you are very risk averse.

I would say that these puzzles are essentially trivial. That's not to say that everyone should find them easy, on the contrary they can be quite baffling. However, mathematically they are trivial, if only because we could find the solutions quickly by simulation, even without a computer. Once we accept that probability is conditional on our knowledge of a system, and learn about Bayes' theorem, they become less surprising.

Intriguing results

In the second category, that of intriguing results, are applications of probability theory that are unexpected or startling. They can not be turned easily into puzzles, because there is too much new content for the puzzle to absorb.

The Birthday Coincidence paradox is comparatively well known and so is not discussed here (you only need 23 people to have better than even odds that at least two of them will have the same birthday). A result that is much less known is Benford's law. This states that in listings and tables of data, the first digit is more likely to be low than high, and in fact the probability that the first digit is n is $\log_{10}(1 + 1/n)$. Benford's law works well in practice, and can be derived from the requirement that any first-digit law must be invariant under a change of scale. It also works however for log tables where the tabulated quantities are dimensionless. Even after seeing a derivation, one is somehow still perplexed by this law. This is probably because of its relative unfamiliarity. Its practical applications are few, unlike the Central Limit theorem, which is perhaps equally perplexing at first sight. The applications by the way, are to such things as detecting fraud that results in made-up data.

A set of interesting paradoxes centres around what is called 'length biased sampling'. When digging up potatoes, why do I usually put my fork through the only really large one? Why does the network connection break during download of the biggest file? These and other perversities of nature result from length-biased sampling: if the connection can break at any time, download of the biggest file takes longer than the others, and so failure is more likely to occur then. Hence the waiting-time paradox: if I arrive at a bus stop at some random time, I am likely to arrive during a long gap between buses, so the next bus takes longer on average to arrive than $\mu/2$, half the mean interarrival time.

We can be more specific about these results, at the cost of introducing some definitions. Recall that, given a probability density

function (pdf) $f(t)$ for occurrence of an event, where t is time, the distribution function F is the probability $F(t) = \int_0^t f(x)dx$ that the event occurs before t , and the survival function $S(t) = 1 - F(t) = \int_t^\infty f(x)dx$ is the probability that the event has not occurred by time t . Later on we shall need the hazard function $h(t)$, where $h(t) dt$ is the probability that the event occurs in the interval $(t, t + dt)$, given that it has not occurred yet, so that $h(t) = f(t) / S(t)$.

With this notation, the distribution of the waiting time has pdf $f_w(t) = S(t) / \mu$, where S is the survival function of the interarrival time distribution. From this, the mean waiting time is $(\mu^2 + \sigma^2) / 2\mu$, where σ^2 is the variance of the interarrival time.

The pdf of time from the last arrival is also $f_w(t)$. It is easy to show that this also applies to ages in a stable population, so that the ages of randomly-met people have pdf $f_w(t)$, where S is now the survival function for the human lifespan. We are more likely to meet people who are around longer.

Biased sampling in general often leads to confusion. Wald found that aircraft returning from combat in WWII had a lower rate of hits near the engines than elsewhere. The military therefore decided that protecting the engines was less important than protecting the rest of the plane. Should we laugh or weep? Let's just hope the story is apocryphal.

I conclude the intriguing results with some paradoxes with practical consequences. The Parrondo paradox is the opposite to length-biased sampling paradoxes, in showing that Nature can sometimes work in our favour. It deserves mention, although it is not specifically a probability paradox. It states that it is possible to win through playing two losing games. For example, both sparrows and insects can eat all the crops, but with a mixture of the two, the sparrows eat the insects and a good crop can be harvested. The Chinese found this out the hard way during the destruction of the 'five pests' in the Cultural Revolution. Simpson's paradox (below) is a probability example of the Parrondo paradox, where the loser in two or more games gets the highest score when they are pooled.

The Condorcet paradox (also called the 'voting paradox') can be stated as that, in games of dice with differently marked 'spinners', it is possible that A can beat B more often than not, B can beat C, but C can beat A. The dice behave like rock-paper-scissors and cannot be ranked in scoring ability like conkers, i.e. 'beating' is not a transitive relationship. An example is quoted in Green [6]. The 'Chinese dice' are marked as follows:

Table 1: Sample dice illustrating Condorcet's paradox. Die A outscores B with probability 5/9, B outscores C with probability 6/9, yet A outscores C with probability of only 3/9.

Die	Markings
Die A	6,6,2,2,2,2
Die B	5,5,5,5,1,1
Die C	4,4,4,3,3,3

Beasley (1989) gives several simple examples of how A can beat B yet B can beat A. The simplest is of a boat race between a sprinting crew and a slower crew with stamina. The sprinting boat can bump the other when it starts in the losing position behind it, and the slow crew will next day eventually catch up with and bump the sprinting crew.

This effect needs to be borne in mind when devising ways of judging sporting events. Perhaps we are too keen to rank people. Human ability can be a many splendoured thing.

Simpson's paradox can be stated as 'what is good for the population as a whole can be bad for every subgroup'. Romano and Siegel (1986) cite 2x2 contingency tables giving recovery rates for people given a drug and for the control group. Both young and old groups did better without the drug, but a higher percentage of the combined population benefited from the drug than did not. Table 2 shows another illustration.

The Ecological Fallacy from epidemiology is that aggregate-level results can not be interpreted at the individual level. Thus, Durkheim found that the suicide rate was greater in regions where a greater percentage of the population was Protestant. There is an obvious explanation, but the data could also be explained if Catholics were more likely to commit suicide in regions where they felt beleaguered.

Table 2: Sample data illustrating Simpson's paradox. A does better than B in both trials, yet B does better overall.

	Successes	Attempts	% Successes
A: first trial	80	100	80.0%
B: first trial	78	100	78.0%
A: second trial	20	40	50.0%
B: second trial	2	5	40.0%
A: total	100	140	71.4%
B: total	80	105	76.2%

The philosophically profound ... or not

'Just because something is infinite, that's no reason to ignore it' (P. A. M. Dirac, speaking at a conference on particle physics in Hungary, 1976 – he was referring to the renormalisation of field theory. I have since then thought of this as 'Dirac's dictum')

There is not a 'martingale paradox', but perhaps there should be (there is a related paradox, the St. Petersburg paradox). The martingale paradox goes as follows: clearly, a player can not systematically win money by betting on the tosses of a fair coin. This remains true even if the player can choose the amounts bet on each toss. Yet a gambling scheme called a martingale (which is banned in casinos) does enable one to win systematically, by doubling the stake after each loss. The player must win a toss eventually, say on the n th toss, and the gain on an original £1 bet is then $2^n - \sum_{j=1}^{n-1} 2^{j-1} = £1$. Hence the paradox.

The 'hidden dragon' is the fact that the *expected* loss before a win is infinite. On average, no pile of money would be large enough to enable me to cover my debts before my eventual win. Yet this explanation is perhaps not completely satisfying, because clearly most of the time I would win within a few tosses. In fact, I was able to surprise a friend by successfully carrying out my announced intention of systematically winning chocolate money off him. What besides indigestion is going on here?

We can see what is happening more clearly if we imagine a stake-doubling strategy where we admit defeat and pay up after n successive losses. The number of losing tosses obeys the geometric distribution, and the probability of winning a pound is $1 - (1/2)^n$. The loss is $\sum_{j=1}^n 2^{j-1} = 2^n - 1$ with probability $(1/2)^n$, so the expected gain is now $1 - (1/2)^n - (2^n - 1)(1/2)^n = 0$, resolving the original paradox. We can not win systematically with a finite pot of money.

As n increases, we lose more and more rarely, but when we do lose, the loss is so great as on average to wipe out all previous

winnings. The martingale strategy has converted a simple 'win some, lose some' gamble into a scheme where we nearly always win a fixed amount, but occasionally lose really badly. As $n \rightarrow \infty$, we have an infinite loss that never occurs! We may wonder in our bleaker moods how many human activities are merely, so to speak, 'martingalised' versions of a fair game. That is, they usually give us a small and predictable gain, but occasionally result in tremendous loss of money or life, so that the average gain from them is zero.

Turning further towards the disquieting, we start to question the foundations of probability, and even of the world itself. Feller's zero probability paradox may shake our concept of probability by showing that events can be possible but yet have probability zero of occurring. Feller gives the example of drawing an exact fraction as a random number from the uniform distribution on $[0,1]$. Although there are infinitely many exact fractions among the real numbers, the probability of drawing one is zero, because, so to speak, there is an infinitely bigger infinity of non-fractional real numbers. The infinitely many fractions occur with probability zero. His proof of this (Feller, 1971, p34) is simple: the ordered sequence $1/2, 1/3, 2/3, 1/4, 3/4, 1/5, 2/5, 3/5, 4/5 \dots$ contains all the rationals in $[0,1]$. Choose $\epsilon < 1/2$ and define an interval ϵ^{k+1} centred on the k th rational. The sum of the interval lengths $\epsilon + \epsilon^2 + \dots = \epsilon / (1 - \epsilon)$ covers the rationals, and has Lebesgue measure zero as $\epsilon \rightarrow 0$. Quite Easily Done!

At this point we might start to wonder what exactly *is* probability. There are several interpretations, none to my mind really satisfactory. The Bayesian–frequentist debate is hardly entertaining, but it does show that even after many years of heated debate, statisticians can not agree on this very fundamental point.

Filled with existential angst, we turn to the Carter/Leslie Doomsday paradox or argument. This has exercised philosophers considerably, and there is currently no consensus. The paradox is discussed at length in Leslie (1998).

The argument is simple. Suppose that the human population is always increasing, let's say exponentially, and will one day fall to zero through some catastrophe. Then most humans of those ever born will live within a few generations of the end of the world. If I think of myself as a 'random' human being, drawn from the number of all who ever have lived or ever will live, then it is most likely that I am one of those living close to the end of the world. Therefore, the end of the world will probably occur soon. This could be good news for those of us hoping to avoid the next research assessment exercise.

This 'anthropic' argument calls to mind the entirely valid length-biased sampling argument discussed earlier. If I don't know where I am in time relative to the arrival of the next bus, I am probably in a long interval between buses rather than a short one. Similarly, is a random person not indeed likely to have 'arrived' within a few generations of the end? On the other hand, it has been argued that a human being and a Martian should calculate the same probability of time to extinction of humanity, and, for the Martian, the anthropic principle would not apply. Further, does it make sense to think of myself as a random person, who arrived in the world through what is rather alarmingly called 'weighted parachuting', and, if we suppose that the End will come through some event unrelated to human activity, how can knowledge of the behaviour of the human population shed any light at all on when that event will arrive? We can begin

to see why the debate has not been settled – this is *really* confusing!

Some simple mathematics is needed if we are to proceed further than the mainly wordy debate has done so far. First, the paradox is stated in a mathematical form. To make the argument concrete, consider an exponentially growing population that grows at rate λ . Denoting time from Doomsday by τ , the proportion $p(\tau)$ of the human race who lived before ‘countdown time’ τ is $p(\tau) = \exp(-\lambda\tau)$. For a random person, p is a random variable from the Uniform distribution, i.e. $p \sim U[0,1]$ and the pdf $f_p(p) = 1$ for $0 < p < 1$. By logic reminiscent of that used to generate random numbers by the inverse method, the pdf $f_d(\tau)$ of τ is given from $f_d(p) dp = f_d(\tau) d\tau$ or

$$f_d(\tau) = \lambda \exp(-\lambda\tau). \quad (1)$$

This is the pdf of time to Doomsday.

What is strange here is that the rate of population increase λ plays the role of the hazard of extinction. If human population growth declines, as is happening at present [8], Doomsday starts to recede. In the real world, population growth may indeed indirectly cause Doomsday, but in this model, which is a mathematical abstraction, extinction can occur entirely as a result of random events, unrelated to human activity, and yet the human growth rate will still appear in equation 1 instead of the (constant) hazard of extinction.

Let us now construct a simple model of extinction of the human population in a more conventional way, by modelling the hazard of extinction p as well as the human population size. The simplest model is of a constant hazard p . This could arise for example from a Poisson process of extinction events such as massive meteor strikes. The pdf of time t to extinction is then

$$f_e(t) = p \exp(-pt), \quad (2)$$

where t is time from the present.

Now if the hazard p is small, the expected time to extinction $1/p$ is large, and vice versa, and the probable time to extinction bears no relation to the growth rate of the human population. Estimation of the distribution of time to Doomsday would be based purely on a study of external threats.

We can here think of ‘probability’ in a frequentist sense, by imagining what would happen in very many computer simulations. This frees us from getting entangled in questions of reasonable prior beliefs. Clearly from this very simple modelling, which does not need to invoke the anthropic principle, we can see that the Carter/Leslie argument must be wrong (phew!), but we have yet to see *why* it is wrong, which is much harder. Note by the way that the argument is not really a paradox by my definition, but rather ‘mathematical flimflam’, of the type discussed by [1].

To see how the argument breaks down, we need to model the human population. Suppose population size is N_0 now (time zero), then at some future time T it will be $N(T) = N_0 \exp(\lambda T)$, and at extinction at time t , $N_f = N_0 \exp(\lambda t)$. Hence, if $F_h(N_f)$ is the distribution function of the final human population N_f ,

$$F_h(N_f) = \text{Prob}(N < N_f) = \text{Prob}(T < t) = 1 - \exp(-pt).$$

Since $t = \log(N_f / N_0) / \lambda$, we have that

$$F_h(N_f) = \exp(-(p/\lambda) \log(N_f / N_0)) = 1 - (N_f / N_0)^{-p/\lambda}. \quad (3)$$

The ultimate size of the human population is thus a random variable from the Pareto distribution. This result is independent of

how the hazard of extinction behaved before time zero, and how the population size of N_0 was reached.

This is a very different result from that given by the Carter/Leslie argument. On any model of population growth before time zero, the proportion of humans who lived before us is small if the hazard p that faces us is small in comparison with the population growth rate λ , whereas if p is much larger than λ , we may expect to be among the last humans to live.

Since, now assuming an infinite past, the proportion of the population who lived before time zero is $p = N_0 / N_f$ we have that

$$F_p(p) = p^{p/\lambda}. \quad (4)$$

To make contact with the Carter/Leslie anthropic argument, we consider the proportion of the population p that precedes a random person living at some epoch, of whom the number is proportional to p . Hence the pdf of the proportion p born before a random person is proportional to $f(p)p$. We must take the limit $p \rightarrow 0$ as otherwise there could not have been an infinite past of exponential growth as in this model. The limiting pdf is then simply a constant, i.e. a person born at any time is indeed equally likely to stand anywhere in the row of all humans who ever lived or will live.

We can thus see that a simple model gives a sensible and intuitive answer, and that the Carter/Leslie argument produces a different result.

I need not weight the probability distribution in equation 4 by p and so take the viewpoint of a ‘random’ person. The existence of many other people similar to myself has no bearing on my investigation into extinction hazards – for that purpose, I might as well be a Martian! The Carter/Leslie equation 1, which would follow from equation 2 on setting $p = \lambda$, has in effect set the hazard of extinction equal to the rate of population growth.

Yet it is still not clear how the Carter/Leslie argument fails. Granted that I do not need to use the anthropic argument, why does it give the wrong result if I *do* use it? It appears to be a sound principle, and the fault lies elsewhere. To see this, having looked into the abyss, we need to tumble right into it and descend into hell. Here the Devil has an interesting paradox reserved for us, again mentioned in Leslie (1998), p235, and due to Eckhardt [4].

The shooting room paradox is this: you are pushed into a room, and told that over 90% of those who ever enter the room will be shot. However, you then learn that the Devil will only shoot the roomful of people if he throws a double six at dice. The odds of this are only 1/36, which is comparatively reassuring. These conflicting odds of death are made possible because people enter the room in batches, each batch being 10 times the size of the previous one. Thus, if the Devil throws a double six at the n th throw, 10^{n-1} people are shot, $\sum_{j=1}^{n-1} 10^{j-1} = (10^{n-1} - 1) / 9$ people are not shot, and the probability of being shot is indeed over 90% ... and (at the same time) 1/36.

What neither Leslie nor the Devil tell us is that this paradox is a disguised version of the martingale paradox discussed earlier. One might call it ‘the Devil’s martingale’. The Devil seeks to increase his gain measured by the number of people he can shoot on ‘winning’ by multiplying the stake by 10 until he does win. The problem for him is that the expected number of people who leave without being shot is infinite. If the Devil admits defeat after n unsuccessful throws, following the discussion of the ‘martingale paradox’ the probability of being shot reverts to 1/36, $\forall n$. Without an infinite supply of people, the Devil can not

guarantee that 90% who enter will be shot, as he may run out of candidates before he throws a double six, and his assertion that at least 90% will be shot means that he is lying. To quote Mandy Rice-Davies 'well, he would, wouldn't he?' Figures 1 and 2 show some simulated realisations of the percentage of those entering the room who are killed, with room-size limits of 10^6 and 10^{30} people respectively, and with the Devil only needing to throw one six. It can be seen that eventually the proportion shot approaches $1/6$.

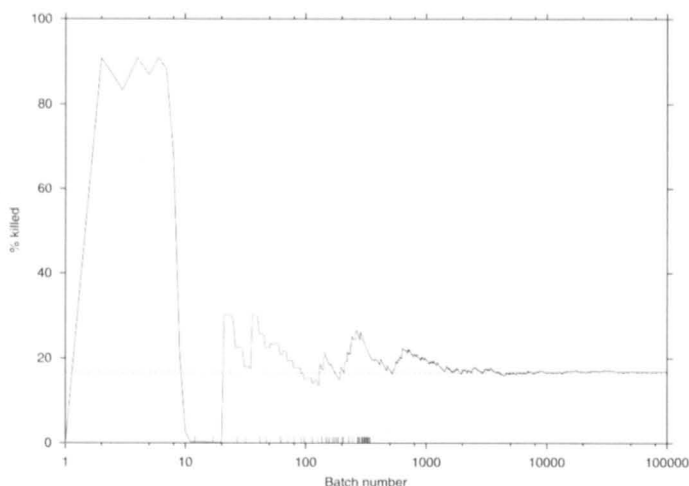


Figure 1 Cumulative percentage of people who enter the 'Devil's shooting room' who are killed, with a room holding 10^6 people and an initial batch of 10 people. The horizontal line is the expected percentage, $100/6$. The small vertical lines show when room size was exceeded and stop after the first 40 for clarity.

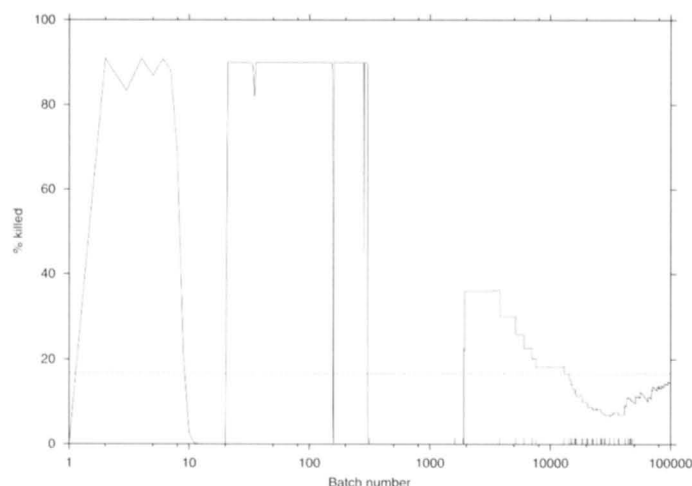


Figure 2 As figure 1, but with a room size of 10^{30} people.

To return briefly to reality, Eckhardt's shooting room paradox was introduced to clarify the debate on the Doomsday paradox. We have the same situation in both cases. In one, a population exponentially expands in continuous time with a constant hazard of extinction. In the other, the scaling up and the extinction take place in discrete time. We can therefore dispose of the Doomsday paradox by regarding it as a martingale problem: to say that the end of the world is nigh is to make the same mistake as to say that one can systematically win at gambling using a martingale. The expected population size at extinction from equation 3 would be infinite if $p < \lambda$. If we allow a population that attains some size N_i to be let off the hook and to not go extinct, then in those cases, so many people survive that the hazard of extinction reduces again

to p , for any finite choice of N_i , in agreement with the simple model of equation 2. This means that if the population is destined to go extinct one day at a level $N_f < N_i$, yes, we are probably near the End. On the other hand, the population we are in is probably one that is destined to reach its target of N_i and not go extinct, because by the anthropic principle itself, most people will be in such large populations. This drops the hazard of extinction back to p . The use of N_i is a mathematical device, and N_i can be taken arbitrarily large (e.g. larger than the number of atoms in the Universe) without affecting the argument. We can even go the limit $N_i \rightarrow \infty$. Approaching the infinite limit in this way avoids ignoring an infinite population gain at infinite time, in accordance with 'Dirac's dictum'.

The Doomsday argument seemed to work because it ignored the infinite number of 'undoomed' people in those realisations of the human population that become very large. It can be regarded as an example of the trickiness of probability in general when concealed infinities creep into the equations, and of martingales in particular. It demonstrates the need for the mathematical rigour of measure theory (e.g. [11]).

The martingale continues to beguile and perplex the human race. It has had a colourful night-life outside the grey pages of books on measure theory (actually, martingales are pretty fascinating even in mathematics, with applications ranging from infectious disease modelling to the stock market). Casanova is known to have successfully used the martingale as a gambling strategy in 1754. Gamblers in 19th century Paris lost a lot of money through it, and since then as we have seen, it has cropped up in a disguised form suited to our anxious age, where it threatens a loss rather than promising a gain. What additional scrapes Cardano might have got into had he known of it!

This concludes this voyage from Monty Hall's Gameshow Heaven to Hell, taking in the End of the World *en route*. My personal conclusion after struggling with some of these paradoxes, is that the human brain is just not wired up to think well probabilistically. Years of exposure to probability theory may make familiar paradoxes seem obvious, but the only way to sort out an unfamiliar one is to switch off one's intuition, go into low gear, and do some mathematics. $\square$

REFERENCES

- 1 Barbeau, E. J. (2000), *Mathematical Fallacies, Flaws, and Flimflam*, The Mathematical Association of America.
- 2 Beasley, J. D. (1989), *The Mathematics of Games*, Oxford University Press, Oxford.
- 3 Carroll, L. (1958), *Pillow Problems and a Tangled Tale*, Dover, New York.
- 4 Eckhardt W. (1997), A Shooting-room View of Doomsday, *The Journal of Philosophy*, **94** (5) 244–259.
- 5 Feller, W. (1971), *An Introduction to Probability Theory and its Applications* vol 2, Wiley, New York.
- 6 Green, D. R. (1981), How probability pays, *Mathematics in School*, **10** (2), 23–24.
- 7 Leslie, J. (1998), *The End of the World: the science and ethics of human extinction*, Routledge, London.
- 8 Lutz, W., Sanderson W. and Scherbov, S. (2001), The end of world population growth, *Nature* **412**, 543–545.
- 9 Ore, O. (1953), *Cardano, the Gambling Scholar*, Dover, London.
- 10 Romano, J. P. and Siegel, A. F. (1986), *Counterexamples in Probability and Statistics*, Wadsworth, Monterey.
- 11 Williams, D. (1991), *Probability with Martingales*, Cambridge University Press, Cambridge.



Rose Baker has a personal chair in Statistics at the University of Salford, where she has worked for the last 25 years. Besides doomsaying, her interests are the applications of Statistics to areas such as finance, epidemiology and healthcare, and maintenance.