expedition to grab the eggs of emperor penguins (*Aptenodytes forsteri*) in midwinter, first detailed by Apsley Cherry-Garrard in *The Worst Journey In the World* (1922). I'm very proud to tell a new story about that well-worn old tale, from my own physical research on site.

And then maybe I'll tell a story about the plan to slow the melting of glaciers in Antarctica by drawing the water out from underneath them. It's now being investigated by a team of glaciologists, as well as governance and finance people, to make sure that it doesn't look like scientists coming in out of left field and telling people how to save the world.

What do you think about AI?

My feeling is that 'artificial intelligence' is a public-relations name that obscures what's really going on. It's artificial for sure. But as for intelligence – the term is so broad that you immediately get lost in it. So AI is a poor name. If it was called 'extremely rapid computation', or 'assisted data analysis' or 'cognitive prosthesis' then that would de-emphasize the magical portions of it. You would be talking about what to do with it, not about making up a human mind or a consciousness.

There's so much bad science fiction that anthropomorphizes AI to the point at which it has agency and malevolence. Machines are not going to get to consciousness using large language models, which is simply optimization. And it's easier to imitate human sentences than we thought it was because we're predictable. So, the Turing test turns out to be a relatively low bar. All you have to do is fool human beings, and we are very gullible.

I had a grand time writing *Aurora* (2015), about a journeying starship, written mainly from the perspective of an AI. The AI, called Ship, represents my thinking on how things might get interesting. It's running a starship and a human says to it: 'keep a narrative account of the trip'. The computer doesn't know what to do and has to figure it out.

It still might not be consciousness, but Ship is pretty eloquent by the end of the novel, pretty self-aware. Pretty much like human consciousness, with the starship as its body and its people like its gut microbiome. But we're talking 500 years from now with a quantum computer. What could happen? Well, one doesn't know.

Do you have a message for scientists?

Scientists need to speak as a group. When all the scientific institutions say, together, 'we, the scientific community, the ones who keep you alive, the ones who are your doctors and provide your food, say this has to be done', that's powerful.

Interview by Anne Pichon.
This interview has been edited for length and clarity.

Does probability exist?

Probably not – but it is useful to act as if it does.
By David Spiegelhalter

Life is uncertain. None of us know what is going to happen. We know little of what has happened in the past, or is happening now outside our immediate experience. Uncertainty has been called the 'conscious awareness of ignorance'[1] – be it of the weather tomorrow, the next Premier League champions, the climate in 2100 or the identity of our ancient ancestors.

In daily life, we generally express uncertainty in words, saying an event "could", "might" or "is likely to" happen (or have happened). But uncertain words can be treacherous. When, in 1961, the newly elected US president John F. Kennedy was informed about a CIA-sponsored plan to invade communist Cuba, he commissioned an appraisal from his military top brass. They concluded that the mission had a 30% chance of success – that is, a 70% chance of failure. In the report that reached the president, this was rendered as "a fair chance". The Bay of Pigs invasion went ahead, and was a fiasco. There are now established scales for converting words of uncertainty into rough numbers. Anyone in the UK intelligence community using the term 'likely', for example, should mean a chance of between 55% and 75% (see go.nature.com/3vhu5zc).

Attempts to put numbers on chance and uncertainty take us into the mathematical realm of probability, which today is used confidently in any number of fields. Open any science journal, for example, and you'll find papers liberally sprinkled with *P* values, confidence intervals and possibly Bayesian posterior distributions, all of which are dependent on probability.

And yet, any numerical probability, I will argue – whether in a scientific paper, as part of weather forecasts, predicting the outcome of a sports competition or quantifying a health risk – is not an objective property of the world, but a construction based on personal or collective judgements and (often doubtful) assumptions. Furthermore, in most circumstances, it is not even estimating some underlying 'true' quantity. Probability, indeed, can only rarely be said to 'exist' at all.

Chance interloper

Probability was a relative latecomer to mathematics. Although people had been gambling with astragali (knucklebones) and dice for millennia, it was not until the French mathematicians Blaise Pascal and Pierre de Fermat started corresponding in the 1650s that any rigorous analysis was made of 'chance' events. Like the release from a pent-up dam, probability has since flooded fields as diverse as finance, astronomy and law – not to mention gambling.

To get a handle on probability's slipperiness, consider how the concept is used in modern weather forecasts. Meteorologists make predictions of temperature, wind speed and quantity of rain, and often also the probability of rain – say 70% for a given time and place. The first three can be compared with their 'true' values; you can go out and measure them. But there is no 'true' probability to compare the last with the forecaster's assessment. There is no 'probability-ometer'. It either rains or it doesn't.

What's more, as emphasized by the philosopher Ian Hacking[2], probability is "Janus-faced": it handles both chance and ignorance. Imagine I flip a coin, and ask you the probability that it will come up heads. You happily say "50–50", or "half", or some other variant. I then flip the coin, take a quick peek, but cover it up, and ask: what's your probability it's heads now?

Note that I say "your" probability, not "the" probability. Most people are now hesitant to give an answer, before grudgingly repeating "50–50". But the event has now happened, and there is no randomness left – just your ignorance. The situation has flipped from 'aleatory' uncertainty, about the future we cannot know, to 'epistemic' uncertainty, about what we currently do not know. Numerical probability is used for both these situations.

There is another lesson in here. Even if there is a statistical model for what should happen, this is always based on subjective assumptions – in the case of a coin flip, that there are two equally likely outcomes. To demonstrate this to audiences, I sometimes use a two-headed coin, showing that even their initial opinion of "50–50" was based on trusting me. This can be rash.

Subjectivity and science

My argument is that any practical use of probability involves subjective judgements. This doesn't mean that I can put any old numbers on my thoughts – I would be proved a poor probability assessor if I claimed with 99.9% certainty that I can fly off my roof, for example. The objective world comes into play when probabilities, and their underlying assumptions, are tested against reality (see 'How ignorant am I?'); but

that doesn't mean the probabilities themselves are objective.

Some assumptions that people use to assess probabilities will have stronger justifications than others. If I have examined a coin carefully before it is flipped, and it lands on a hard surface and bounces chaotically, I will feel more justified with my 50–50 judgement than if some shady character pulls out a coin and gives it a few desultory turns. But these same strictures apply anywhere that probabilities are used — including in scientific contexts, in which we might be more naturally convinced of their supposed objectivity.

Here's an example of genuine scientific, and public, importance. Soon after the start of the COVID-19 pandemic, the RECOVERY trials started to test therapies in people hospitalized with the disease in the United Kingdom. In one experiment, more than 6,000 people were randomly allocated to receive either the standard care given in the hospital they were in, or that care plus a dose of dexamethasone, an inexpensive steroid[3]. Among those on mechanical ventilation, the age-adjusted daily mortality risk was 29% lower in the group allocated dexamethasone compared with the group that received only standard care (95% confidence interval of 19–49%). The *P* value — the calculated probability of observing such an extreme relative risk, assuming a null hypothesis of no underlying difference in risk — can be calculated to be 0.0001, or 0.01%.

This is all standard analysis. But the precise confidence level and *P* value rely on more than just assuming the null hypothesis. It also depends on all of the assumptions in the sta-

"Probability is 'Janus-faced': it handles both chance and ignorance."

tistical model, such as the observations being independent: that there are no factors that cause people treated more closely in space and time to have more-similar outcomes. But there are many such factors, whether it's the hospital in which people are being treated or changing care regimes. The precise value also relies on all of the participants in each group having the same underlying probability of surviving 28 days. This will differ for all sorts of reasons.

None of these false assumptions necessarily mean that the analysis is flawed. In this case, the signal is so strong that a model allowing, say, the underlying risk to vary between participants will make little difference to the overall conclusions. If the results were more marginal, however, it would be appropriate to do extensive analysis of the model's sensitivity to alternative assumptions.

To exercise the much-quoted aphorism, "all models are wrong, but some are useful"[4]. The dexamethasone analysis was particularly useful because its firm conclusion changed clinical practice and saved hundreds of thousands of lives. But the probabilities that the conclusion was based on were not 'true' — they were a product of subjective, if reasonable, assumptions and judgements.

Down the rabbit hole

But are these numbers, then, our subjective, perhaps flawed estimates of some underlying 'true' probability, an objective feature of the world?

I will add the caveat here that I am not talking about the quantum world. At the sub-atomic level, the mathematics indicates that causeless events can happen with fixed probabilities (although at least one interpretation states that even those probabilities express a

How ignorant am I?

The need to evaluate the accuracy of the probabilities we assign to things became clear when weather forecasters started giving probabilities of precipitation. In 1951, meteorologist Glenn Brier developed the Brier score as a way to assess predictions[12], and it can be adapted to see how good or bad you are at assessing your degree of confidence about facts.

A good Brier score depends on a probability assessor both being discriminatory, so they give some confident judgements, but also calibrated, so that of the situations in which they state '70% probability', they are right around 70% of the time. This idea turns out to be fundamentally important when we consider the meaning of these subjective judgements (see main text).

For each of the following questions, decide which answer you feel is most likely to be correct, and then quantify your confidence on a scale from 5 to 10. For example, if you are certain that answer (A) is correct, you should give it 10/10, but if you are only around 70% sure, then it gets 7/10. If you have no idea, then give 5/10 to either choice.

Questions

1. Which contains more water by percentage?
   - (A) Human brain
   - (B) Human blood

2. Who published their key work first?
   - (A) Charles Darwin
   - (B) Gregor Mendel

3. Which is the bigger planet?
   - (A) Venus
   - (B) Earth

4. Which molecule has more atoms?
   - (A) Caffeine
   - (B) Aspirin

5. Which has the higher melting point?
   - (A) Gold
   - (B) Silver

Now, check your judgements against the answers on the facing page (also do this quiz online at go.nature.com/3vakwqj). The following table shows how you should score yourself when the true answer is revealed.

Your confidence that your answer is correct (out of 10):	5	6	7	8	9	10
Score if you are right:	0	9	16	21	24	25
Score if you are wrong:	0	−11	−24	−39	−56	−75

The scoring is deliberately harsh (it uses the square of the prediction error). By punishing failure more than rewarding success, honesty is encouraged.

If you ended up with a negative total, you did worse than a complete ignoramus who just answered 5 to every question. People with an exaggerated sense of their own knowledge tend to end up with large negative scores. Those with an awareness of their own doubts tend to mainly use 5s, 6s or 7s, and might end up with a small positive score. People who actually know a lot, or are extremely lucky, get higher scores. This type of exercise is used to train forecasters to be less over-confident, and have insight into their own thought processes.

relationship with other objects or observers, rather than being intrinsic properties of quantum objects)[5]. But equally, it seems that this has negligible influence on everyday observable events in the macroscopic world.

I can also avoid the centuries-old arguments about whether the world, at a non-quantum level, is essentially deterministic, and whether we have free will to influence the course of events. Whatever the answers, we would still need to define what an objective probability actually is.

Many attempts have been made to do this over the years, but they all seem either flawed or limited. These include frequentist probability, an approach that defines the theoretical proportion of events that would be seen in infinitely many repetitions of essentially identical situations — for example, repeating the same clinical trial in the same population with the same conditions over and over again, like Groundhog Day. This seems rather unrealistic. The UK statistician Ronald Fisher suggested thinking of a unique data set as a sample from a hypothetical infinite population, but this seems to be more of a thought experiment than an objective reality. Or there's the semi-mystical idea of propensity, that there is some true underlying tendency for a specific event to occur in a particular context, such as my having a heart attack in the next ten years.

This seems practically unverifiable.

There is a limited range of well-controlled, repeatable situations of such immense complexity that, even if they are essentially deterministic, fit the frequentist paradigm by having a probability distribution with predictable properties in the long run. These include standard randomizing devices, such as roulette wheels, shuffled cards, spun coins, thrown dice and lottery balls, as well as pseudo-random number generators, which rely on non-linear, chaotic algorithms to give

"Any practical use of probability involves subjective judgements."

numbers that pass tests of randomness.

In the natural world, we can throw in the workings of large collections of gas molecules which, even if following Newtonian physics, obey the laws of statistical mechanics; and genetics, in which the huge complexity of chromosomal selection and recombination gives rise to stable rates of inheritance. It might be reasonable in these limited circumstances to assume a pseudo-objective probability — 'the' probability, rather than 'a' (subjective) probability.

In every other situation in which probabilities are used, however — from broad swathes of science to sports, economics, weather, climate, risk analysis, catastrophe models and so on — it does not make sense to think of our judgements as being estimates of 'true' probabilities. These are just situations in which we can attempt to express our personal or collective uncertainty in terms of probabilities, on the basis of our knowledge and judgement.

Matters of judgement

This all just raises more questions. How do we define subjective probability? And why are the laws of probability reasonable, if they are based on stuff we essentially make up? This has been discussed in the academic literature for almost a century, again with no universally agreed outcome.

One of the first attempts was made in 1926 by the mathematician Frank Ramsey at the University of Cambridge, UK. He ranks as the person in history I would most like to meet. He was a genius whose work in probability, mathematics and economics is still considered fundamental. He worked only in the mornings, devoting his after-hours to a wife and a lover, playing tennis, drinking and enjoying exuberant parties while laughing "like a hippopotamus" (he was a big man, weighing in at

108 kilograms). He died in 1930 aged just 26, probably, according to his biographer Cheryl Misak, from contracting leptospirosis after swimming in the River Cam[6].

Ramsey showed[7] that all the laws of probability could be derived from expressed preferences for specific gambles. Outcomes have assigned utilities, and the value of gambling on something is summarized by its expected utility, which itself is governed by subjective numbers expressing partial belief — that is, our personal probabilities. This interpretation does, however, require an extra specification of these utility values. More recently, it's been shown[8] that the laws of probability can be derived simply by acting in such a way as to maximize your expected performance when using a proper scoring rule, such as the one shown in the text box "How ignorant am I?".

Attempts to define probability are often rather ambiguous. In his 1941–2 paper 'The Applications of Probability to Cryptography', for example, Alan Turing uses the working definition that "the probability of an event on certain evidence is the proportion of cases in which that event may be expected to happen given that evidence"[9]. This acknowledges that practical probabilities will be based on expectations — human judgements. But by "cases", does Turing mean instances of the same observation, or of the same judgements?

The latter has something in common with frequentist definition of objective probability, just with the class of repeated similar observations replaced by a class of repeated similar subjective judgements. In this view, if the probability of rain is judged to be 70%, this places it in the set of occasions in which the forecaster assigns a 70% probability. The event itself is expected to occur in 70% of such occasions. This is probably my favourite definition. But the ambiguity of probability is starkly demonstrated by the fact that, after nearly four centuries, there are many people who won't agree with me on that.

Pragmatic approach

When I was a student in the 1970s, my mentor, statistician Adrian Smith, was translating the Italian actuary Bruno de Finetti's *Theory of Probability*[10]. De Finetti had developed ideas of subjective probability at around the same time as Ramsey, but entirely independently. (They were very different characters: in contrast to Ramsey's staunch socialism, in his youth de Finetti was an enthusiastic supporter of Italian dictator Benito Mussolini's style of fascism, although he later changed his mind.) That book begins with the provocative statement: "probability does not exist", an idea that has had a profound influence on my thinking over the past 50 years.

In practice, however, we perhaps don't have to decide whether objective 'chances' really exist in the everyday non-quantum world. We

Imprecise expression of uncertainty helped to persuade John F. Kennedy to back a CIA-organized invasion of Cuba.

can instead take a pragmatic approach. Rather ironically, de Finetti himself provided the most persuasive argument for this approach in his 1931 work on 'exchangeability', which resulted in a famous theorem that bears his name[11]. A sequence of events is judged to be exchangeable if our subjective probability for each sequence is unaffected by the order of our observations. De Finetti brilliantly proved that this assumption is mathematically equivalent to acting as if the events are independent, each with some true underlying 'chance' of occurring, and that our uncertainty about that unknown chance is expressed by a subjective, epistemic probability distribution. This is remarkable: it shows that, starting from a specific, but purely subjective, expression of convictions, we should act as if events were driven by objective chances.

It is extraordinary that such an important body of work, underlying all of statistical science and much other scientific and economic activity, has arisen from such an elusive idea. And so I will conclude with my own aphorism. In our everyday world, probability probably does not exist — but it is often useful to act as if it does.

David Spiegelhalter is emeritus professor of statistics at the University of Cambridge, UK. He is the author of *The Art of Uncertainty: How to Navigate Chance, Ignorance, Risk and Luck* (Penguin, 2024).

1. Smithson, M. *Ignorance and Uncertainty: Emerging Paradigms* (Springer, 1989).
2. Hacking, I. *The Emergence of Probability* (Cambridge Univ. Press, 2006).
3. The RECOVERY Collaborative Group *N. Engl. J. Med.* **384**, 693–704 (2021).
4. Box, J. E. P. *J. Am. Stat. Assoc.* **71**, 791–799 (1976).
5. Rovelli, C. *Helgoland: The Strange and Beautiful Story of Quantum Physics* (Penguin, 2022).
6. Misak, C. *Frank Ramsey: A Sheer Excess of Powers* (Oxford Univ. Press, 2020).
7. Ramsey, F. P. in *The Foundations of Mathematics and other Logical Essays* 156–198 (Routledge, 1926).
8. Lindley, D. V. *Int. Stat. Rev.* **50**, 1–11 (1982).
9. Turing, A. M. UK National Archives record HW 25/37 (1941–2).
10. de Finetti, B. *Theory of Probability* (Wiley, 1974).
11. de Finetti, B. *Atti della R. Accad. Naz. Lincei* **4**, 251–299 (1931).
12. Brier, G. W. *Mon. Weather Rev.* **78**, 1–3 (1950).

Answers to the quiz

1. A   (The brain is about 73% water, blood is about 50% water)

2. A   (Darwin published *On the Origin of Species* in 1859, Mendel published *Experiments on Plant Hybridization* in 1866)

3. B   (Venus's radius is 6,052 kilometres, Earth's is 6,378 kilometres)

4. A   (Caffeine has 24 atoms, aspirin 21 atoms)

5. A   (Gold's melting point is 1,064 °C, silver's is 962 °C)

MICHAEL OCHS ARCHIVES/GETTY

Correction

The picture caption in this story did not accurately capture the circumstances surrounding the Cuba invasion.