

IN THE YEAR 2054: INNUMERACY DEFEATED

GERD GIGERENZER

It is the year 2054. Our great-granddaughters and great-grandsons are celebrating a triple anniversary: the 400th anniversary of the mathematical theory of probability, the 200th of George Boole's *The Laws of Thought*, and the 100th of the publication of Leonard Savage's *Foundations of Statistics*. This year's celebration happens to coincide with the final victory over an intellectual disability that has plagued humankind for centuries: innumeracy, or the inability to think with numbers, specifically numbers that represent uncertainties and risks. Where are we? In Paris. It was in France that probability theory was born, back in the seventeenth century. The Great Hall at the Sorbonne is packed with flowers and guests, and is presided over jointly by the president of France and the president of the World Health Organization. A large curved podium provides the set for the four most distinguished scholars in the social sciences. At least, this is what the programme says. Other scholars in the audience think they should have been asked to speak, but Fortuna was not with them. The topic of this afternoon's panel discussion is 'How the war against innumeracy was won'. The chair is Professor Emile Ecu, an economist at the Sorbonne.

CHAIR: Madame le President, Monsieur le President, dear panel, guests, and audience. We have exactly 30 minutes to reconstruct what is arguably the greatest success of the social sciences in the twenty-first century, the defeat of innumeracy. The twentieth century had eradicated illiteracy, that is, the inability to read and write, at least in France. The challenge to our century was innumeracy. The costs of innumeracy have been a tremendous financial burden to modern economies, as had been those of illiteracy before. This year, the war against innumeracy has been declared won by the World Health Organization. Let us ask our distinguished panelists how this success came about?

SOCIOLOGIST (Paris): It all began with a programmatic statement by the father of modern science fiction, Herbert George Wells, best known, perhaps, as the author of *The Time Machine*. At the beginning of the twentieth century, Wells predicted that 'statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write'. His message spread through all of the influential works that eventually led to the eradication of innumeracy. Here, we have a wonderful case in which literature eventually incited a revolution.

STATISTICIAN (Beijing): Wells never said that. That quotation was made up, most likely by Darrel Huff, who used it as an epigraph in his bestseller, *How to lie with statistics*. Making up quotations was consistent with the title of his book. Always check your sources!

CHAIR: Are you saying that the crusade against innumeracy was started by a fake?

SOCIOLOGIST: No, this is not . . .

STATISTICIAN: Yes, the Wells quote was a fake, and it did not initiate the crusade against innumeracy. It wasn't modern literature. It all started exactly 400 years ago with one of the greatest intellectual revolutions, the probabilistic revolution. At this time a notion of rationality was developed that eventually replaced the old ideal of certainty with a new, modest conception of rational belief that acknowledged uncertainty. The initiators of this revolution were the mathematicians Blaise Pascal and Pierre Fermat who, between July and October of 1654, exchanged letters about problems posed by a notorious gambler and man about town, the Chevalier de Méré. This revolution was not just an intellectual one—the calculus of probability—it was also a moral one, the . . .

CHAIR: What's moral about gambling?

STATISTICIAN: Not gambling, probability. Take Pascal's wager. Before Pascal, people believed in God because they were absolutely certain that He existed. For Pascal, God existed only as a probability, and the decision to believe or not to believe in God should be the outcome of a rational calculation—not of blind faith or stubborn atheism. According to Pascal's reasoning, you could make two errors. If God does not exist but you nevertheless believe, you might forgo some worldly pleasures. That's bad news. But if He does exist and you do not believe, then you face eternal damnation. That's not just bad news; that's a never-ending disaster. Thus, even if there is only a small probability that God exists, the pay-off is infinite—infinite bliss for those who believe and infinite misery for the others who, mistakenly, do not. Here is the beauty of Pascal's wager: The meaning of being moral had changed from blind faith and illusory certainty to rational self-interest and cost-benefit calculations.

CHAIR: I always knew that God was an economist. She must have loved this Pascal. I believe it was he who said, 'The heart has its reasons which reason knows nothing of'.

PSYCHOLOGIST (Boston): Let's get back down to earth. Whatever the moral implications of the calculus of probability, it was soon discovered that ordinary minds didn't understand probabilities, at least most of the time. This phenomenon was called innumeracy.

STATISTICIAN: No, no, no; not so fast. On the contrary, Pierre Laplace and, even earlier, the Enlightenment mathematicians said that probability theory is just common sense reduced to a calculus, and that educated persons—*les hommes éclairés*—have this common sense. And so did George Boole in 1854, when he set out to derive the laws of probability and logic from the laws of thought. It was known that people were occasionally confused by probabilities; Laplace himself described the 'gambler's fallacy' and other errors. But these mistakes were thought to result from the intervention of emotion and wishful thinking into rational processes. The two major rules of probability were definitely believed to be descriptions of actual human reasoning: the law of large numbers by Jacob Bernoulli, and the rule of inverse probabilities by Thomas Bayes.

HISTORIAN OF SCIENCE (Cambridge): Be that as it may, Bayes' rule is not from Bayes. He seems to have copied it from Nicholas Saunderson, who held the most prestigious academic chair in England, the Lucasian chair of Mathematics at Cambridge, which

Newton had held before. My dear friend, you should know Stigler's Law of Eponymy that says that no scientific discovery is named after its original discoverer.¹

CHAIR: Hmm . . . I wonder who discovered Stigler's law . . .

HISTORIAN OF SCIENCE: One impolite reading of this law is that every scientific discovery is named after the last individual too ungenerous to give due credit to his predecessors.

STATISTICIAN: Why would that point be relevant here?

HISTORIAN OF SCIENCE: A minute ago, you yourself cared about the origins of a quotation attributed to Wells and I became concerned about the origins of a rule attributed to Bayes. Who cares? Pascal's triangle is not from Pascal, Gauss' law is also not from Gauss, and the Pythagorean theorem is not . . .

PSYCHOLOGIST: Can we move on? Innumeracy was described by the mathematician John Allen Paulos, who, in 1988, wrote a bestseller of the same title. For instance, he related the story of a weather forecaster on American television who reported that there was a 50% chance of rain on Saturday and a 50% chance of rain on Sunday, and then concluded that there was a 100% chance of rain that weekend.

SOCIOLOGIST: Paulos did not discover innumeracy. That phenomenon had already been described in the 1970s and 1980s by psychologists, notably Daniel Kahneman and Amos Tversky. In a series of experiments, they showed that people are confused by probabilities. They did not talk about innumeracy, but about 'cognitive illusions'. But it's the same thing: the base rate fallacy, the conjunction fallacy, the . . .

HISTORIAN OF SCIENCE: The conjunction fallacy had actually been described by Bärbel Inhelder and Jean Piaget in 1958 in their book on the early growth of logic in the child; they just used a different term, namely set inclusion rather than conjunction.² It's the same phenomenon. The base rate fallacy was actually discovered by the French mathematician . . .³

CHAIR: I thought that historians had already given up priority questions in the last century. What's the point of priority if there is no patent, copyright, or other source of income at stake?

PSYCHOLOGIST: OK, by the 1980s, there was ample evidence for a phenomenon eventually labelled innumeracy, the inability to think about uncertainties and risks.

CHAIR: Why was innumeracy diagnosed so late? I have heard that our Statistician and our Sociologist have different opinions on this question.

SOCIOLOGIST: May I go first? In the 1960s, the Western world was confronted with a flood of seemingly irrational behaviour, from the assassinations of John F. Kennedy and Martin Luther King, Jr. in the United States to the violent student revolutions of 1968 in countries all over the world. These events shattered the ideal of reasonable discourse

¹ Stigler, S. M. (1983). Who discovered Bayes' Theorem? *American Statistician*, 37:4, 290–296.

² Inhelder, B. & Piaget, J. (1958). *Growth of logical thinking: From childhood to adolescence*. New York: Basic Books.

³ The historian of science seems to refer to Rouanet, H. (1961). Études de décisions expérimentales et calcul de probabilités. [Studies of experimental decision making and the probability calculus]. In *Colloques Internationaux du Centre National de la Recherche Scientifique* (pp. 33–43). Paris, France: Éditions du Centre National de la Recherche Scientifique.

and brought human irrationality to the foreground. Something similar had happened during the French Revolution, the bloody aftermath of which destroyed the idea that common sense would follow the calculus of reason, as probability theory was then called . . .

CHAIR: By the way, if I may interrupt, the events following the French Revolution were one reason why the subjective interpretation of probability—the idea that the laws of probability are about reasonable degrees of belief—was discredited by 1830 and the frequency interpretation began to reign. Degrees of belief came to be thought of as too disorderly to be a proper subject matter for the theory of probability, as opposed to orderly frequencies of things like mortality, suicide, prostitution . . .

SOCIOLOGIST: Fine, but my thesis is that some degree of innumeracy had always been a facet of human minds; it was simply amplified by the political events of the 1960s, as it had been before by those of the French Revolution. Psychologists of the 1970s just took advantage of the political climate and claimed that human disasters of any kind, including racial prejudice and ‘hot’ social behaviour, could be explained by ‘cold’ cognitive illusions. Consistent with my hypothesis, errors in statistical reasoning were no big deal in psychology before the political events of the 1960s.

STATISTICIAN: I think that looking for political causes for innumeracy is a bit too far-fetched. Neither the assassinations in the United States nor the political turmoil of the cultural revolution in China, if I may add another event of the 1960s, produced innumeracy. It did not come from outside influences, it came from within the decision theorists themselves. Listen carefully: whether you like it or not, innumeracy was, to some degree, created by decision theorists like Leonard Savage . . .

(Unrest in the audience.)

STATISTICIAN: . . . with their ultra-liberal, one might even say, expansionist policy of extending the laws of probability to everything between heaven and earth. These neo-Bayesians were not satisfied that probabilities mean observable frequencies; no, they claimed that one can and should attach a probability to everything. This created massive confusion in ordinary minds. Please recall what kind of statements confused people. These were statements involving probabilities, specifically single-event probabilities and conditional probabilities! Savage popularized single-event probabilities and Ward Edwards and others brought the message to the social sciences. For our Historian of Science, I add that Savage built on von Neumann and Morgenstern’s work in the 1940s, so he does not have to lecture us on that.

(Unrest in the audience finds relief in laughter.)

STATISTICIAN: Since about 1830, probability has been interpreted as a relative frequency in a reference class, or sometimes as a physical propensity, and this gave the laws of probability a well-defined, although modest, realm of application. Unlike many of his followers, Savage was aware of the oddity of his proposal.

CHAIR: You mean extending the laws of probability to messy mental products such as degrees of belief?

STATISTICIAN: Exactly. In his 1954 book, Savage began his chapter on personal probability by saying that he considers it more probable that a Republican president will be elected in 1996 than that it will snow in Chicago sometime in the month of May 1994. And then he added that many people, after careful consideration, are convinced that such subjective probabilities mean precisely nothing, or at best, nothing precisely.

PSYCHOLOGIST: Right! We actually had a Democratic, not a Republican president in 1996 and there was no snow in Chicago in May 1994. So Savage was wrong . . .

SOCIOLOGIST: No, he wasn’t. A probability statement about a single event can never be wrong, except when the probabilities are 0 or 1.

HISTORIAN OF SCIENCE: Look, our Statistician’s thesis is that the extension of the laws of probability to degrees of belief, including beliefs about singular events, confused people and provided fertile ground for demonstrating reasoning fallacies. In other words, without this extension, there would not have been that magnitude of innumeracy. What she means is that most reasoning fallacies were demonstrated with probabilities rather than with frequencies.

SOCIOLOGIST: Don’t be blind to the political dimension! How was it that before 1968 almost all psychologists agreed that man is a good intuitive statistician—pardon the sexist language of those days—and only a few years later, from the 1970s on, the same people embraced the opposite message? That change was not supported by fact, but rather grew out of a new political climate in which irrationality got the applause, and if one looked long enough . . .

CHAIR: Here is the disagreement: our Sociologist argues that the interest in statistical innumeracy was merely amplified by the political events of the 1960s. Our Statistician offers the conjecture that innumeracy itself had already been partly created in 1954 by Savage’s extension of the laws of probability beyond frequencies, specifically to singular events. Ordinary people, she assumes, are frequentists.

HISTORIAN OF SCIENCE: Neither of these two interpretations—political turmoil or over-extension of the meaning of mathematical probability—was ever discussed in the 1970s. The explanation usually presented was that people simply suffer from cognitive illusions just as they suffer from visual illusions. Our Statistician’s hypothesis should not be misread in the sense that innumeracy was not real; that would be a misunderstanding. The newspapers and the medical textbooks began to use single-event probabilities and conditional probabilities, and citizens and students alike were confused, often without even noticing it. That was all real. There were even court trials over the meaning of single-event probabilities. In one case, the prosecution had offered a defendant a plea bargain, but his lawyer told him that he had a 95% chance of acquittal based on an insanity plea. Based on this probability, the defendant rejected the plea bargain, stood trial, and was sentenced to 20 years in prison for first-degree murder. So he sued his attorney for having given him an unrealistic probability. Courts had to deal with the question, ‘Can a single-event probability be wrong?’ If it cannot, then what does it mean?

STATISTICIAN: The same confusion emerged in everyday life when institutions started to communicate all kinds of uncertainties in probabilities. In 1965, the US National Weather Service began to express forecasts in probabilities . . .

SOCIOLOGIST: That quantophobia never occurred in France!

STATISTICIAN: . . . such as that there is a 30% chance of rain tomorrow. Most Americans thought they knew what that meant. However, studies showed that some people understood this statement to mean that it will rain in 30% of the area, others that it will rain 30% of the time tomorrow, and others that it will rain on 30% of the days like tomorrow. A single-event probability leaves, by definition, the reference class open: area, hours, days, or something else. But people, then and now, think in terms of concrete cases and fill a class in.

PSYCHOLOGIST: The insight that people tend to construct reference classes became the basis of the first theory, in the early 1990s, that linked the cognitive processes underlying judgements of confidence with those underlying judgements of frequency; this theory, in turn, revealed how to make a celebrated case of innumeracy, the overconfidence bias, appear and disappear. This was first shown in laboratory research with general-knowledge questions, in which participants were asked, 'What is the probability that your answer is correct?' and then, after a number of questions, 'How many of your answers are correct?' Judgments of probability were systematically higher than judgements of frequencies. Subsequently, the same result was shown for highly consequential decisions such as whether or not a prisoner should be given conditional freedom, that is, probation, bail, or weekend leave. In one influential study, a group of probation officers was asked 'What is the probability that Mr. Smith will commit a violent act if he is discharged?' whereas another was asked 'Think of 100 men like Mr. Smith. How many of them will commit a violent act if they are discharged?' When the average estimate of the probability of harm was 0.30, the frequency estimate was only 20 of 100.⁴

CHAIR: How so?

PSYCHOLOGIST: For the same reason probability judgements were higher in estimations of general knowledge and weather forecasting. Asking for a single-event probability leaves the reference class open. The probation officers themselves need to fill one in: does the probability refer to the situation that Mr. Smith is on weekend release 100 times, or that 100 people like Mr. Smith are on weekend release once, or something else? The answer need not be the same. The frequency question specifies a reference class, the probability question leaves it open and thus can lead to systematic differences in the answers.

CHAIR: If I recall correctly, the idea was initially not well understood that reference classes, as proposed by probabilistic mental models theory, could be at the core of the overconfidence bias, as well as the systematic difference between judgements of single events and frequencies. Many understood that frequency judgements are always right, period; they missed the real issue, namely changing reference classes, which allowed the prediction of when frequency judgements were right or wrong. Some even thought that the difference occurs because frequencies are more frightening or elicit more emotional responses.

STATISTICIAN: In the 1990s, *three* reasons were found to explain why people were confused by probabilities. Unspecified reference classes in single-event statements was the one you mentioned; polysemy and computational complexity were the other two. The three together explained much of the confusion, not all of it, and also provided the key to overcoming it.

⁴ On the laboratory research see Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98:506–528. On the research with experts in law and psychiatry see Slovic, P., Monahan, J. & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, 24:271–296.

CHAIR: That's what I meant; often all three explanations were jumbled together as the 'frequency effect'.

PSYCHOLOGIST: I don't think that polysemy is as interesting a reason as the other two. That part was just a game with words with multiple meanings to produce reasoning that looked like innumeracy.

CHAIR: What do you mean?

PSYCHOLOGIST: The game went like this. Experimental participants read a description of a woman called Linda that made her look like a feminist, and then they were asked whether it is more probable that she is a bank teller, or a bank teller and active in the feminist movement; most chose the latter. That was called a conjunction fallacy and was interpreted as demonstrating that the human mind is not built to work with the laws of probability. But when people were asked a frequency question, 'Think of 100 women like Linda. How many are bank tellers? How many are bank tellers and active in the feminist movement?', that alleged fallacy disappeared. Only at that point did it become clear that the problem was with the polysemy of the term probability, that is, with its several legitimate non-mathematical meanings, including 'plausible', 'typical', and 'whether there is evidence'. A frequency question eliminated this polysemy and elicited a different answer. German psychologists were the first to show that the so-called conjunction fallacy disappeared with frequency questions . . .⁵

HISTORIAN OF SCIENCE: No, it was Bärbel Inhelder and Jean Piaget who had shown long before that children can understand conjunctions; they just used the term set inclusions. They put a box containing wooden beads, most of them were brown but two were white, in front of a child and asked: 'Are there more wooden beads or more brown beads in this box?' By the age of eight, a majority of the children responded that there were more wooden beads in the box. Note that Piaget and Inhelder had asked children about frequencies, not probabilities.

CHAIR: At that time, didn't anybody ask why Stanford University students should suffer from the conjunction fallacy when Genevieve children, by the age of eight, didn't?

PSYCHOLOGIST: Few researchers asked any questions. At that time, most of them were busy running experiments; that was during the empiricist phase, before psychology finally became a full-blown theoretical science in 2010. There were even scholars who claimed that human disasters of many kinds, including US foreign policy in the late twentieth century, could be explained by the conjunction fallacy;⁶ there were scholars who . . .

SOCIOLOGIST: Yeah, yeah, yeah, and there were statesmen who claimed that AIDS was caused by poverty, not by HIV. But I agree that, unlike the other two reasons for public confusion with probability, polysemy had no real practical consequences.

⁵ The psychologist refers to Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50:123–129; and Hertwig, R. & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12:275–305. The historian of science refers to Inhelder, B. & Piaget, J. (1958). *Growth of logical thinking: From childhood to adolescence*. New York: Basic Books.

⁶ Kanwisher, N. (1989). Cognitive heuristics and American security policy. *Journal of Conflict Resolution*, 33:652–675.

HISTORIAN OF SCIENCE: Uncovering the effect of external representations on statistical thinking started the ball rolling against innumeracy. That is, it was realized that the problem was not just a lack of training in the laws of statistics; innumeracy could be tackled in a much easier way by realizing that some representations of uncertainty help people understand uncertainties, whereas others do not. Unfortunately, the unhelpful representations were the ones generally used in teaching, medicine, by the media, in the court of law . . .

STATISTICIAN: Consistent with my thesis, if I might add, among the representations which tend to confuse people were those promoted by Savage, such as the use of single-event probabilities . . .

HISTORIAN OF SCIENCE: Yes, but also others, such as conditional probabilities and relative risks. In contrast, representations that were found to foster insight included those that specify reference classes, make use of natural frequencies, and communicate absolute risks. Teaching people to use proper representations turned out to be a fast and effective method against innumeracy. Natural frequencies, for example, are much more effective than conditional probabilities; as mentioned earlier, they reduce the computational complexity in making inferences. This was discovered in the mid-1990s by the same two German psychologists who had earlier worked out the crucial role of reference classes in their probabilistic mental models theory, namely . . .

PSYCHOLOGIST: No, no; before that, in 1982, two American scholars, Christensen-Szalanski and Beach, had used natural sampling, of which natural frequencies are the result.⁷ Coming under attack by those who did not want to see cognitive illusions disappear, they did not develop the idea further and did not have any influence; but they were first.

HISTORIAN OF SCIENCE: Whoever was first, the insight was that natural frequencies facilitate Bayesian reasoning; that is, they facilitate the estimation of a posterior probability (or a frequency) from observation.⁸ This insight turned the question of innumeracy into an ecological, perhaps even an evolutionary, one.

CHAIR: Bayes and evolution?

PSYCHOLOGIST: The argument was that animals and humans have spent most of their evolution in an environment of natural sampling—before the development of probability theory or statistical surveys. Natural frequencies are the result of natural sampling. For instance, take a physician in an illiterate society who is confronted with a new disease. The physician has observed 20 cases of a symptom with the disease and 40 cases of the symptom without the disease. When a new case with the symptom comes in, she can easily compute the Bayesian posterior probability of this patient having the disease: $20/(20 + 40)$, which is one-third. That's how Bayesian inference was done before 1654, that is, before mathematical probabilities were introduced.

⁷ Christensen-Szalanski, J. J. & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior and Human Performance*, 29:270–278.

⁸ Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102:684–704; Gigerenzer, G. & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: A reply to Lewis & Keren and Mellers & McGraw. *Psychological Review*, 106:425–430.

CHAIR: Oh, now I understand the paradox that animals were reported to be good Bayesians but humans were not. It's the representation, not the species. Animals encode natural frequencies, and we poor humans got conditional probabilities . . .

PSYCHOLOGIST: Right. Natural frequencies were transformed into conditional probabilities; for instance, dividing the 20 cases of the symptom with the disease by the total number of disease cases, say 25, results in a probability $p(\text{symptom}|\text{disease})$ of .80. The physician in the illiterate society can ignore the base rate of 25—it is not necessary for computing the posterior probability, as we just saw. Not so the medical student who has to multiply the conditional probabilities by the respective base rates in order to get the base rate information back—which amounts to the rather complicated form in which Bayes' rule had been taught, and which most physicians never understood. Thus, the insight was that Bayesian computations depend on the representation of the information in the environment—that's the ecological part. The second insight was that the cure is to use a representation that humans had encountered during most of their evolutionary history, natural frequencies, that is, frequencies that have not yet been normalized with respect to the base rates—that's the historical, or evolutionary, part.

CHAIR: Was this the point where the World Health Organization took over? In educating physicians to understand uncertainties?

HISTORIAN OF SCIENCE: No, this laboratory research had its impact first in the Anglo-Saxon courts, possibly because of their adversarial procedure. Defence lawyers realized that confusions due to probabilities were not in their or their clients' interests, but typically *were* in the prosecution's interests. For many decades, experts had testified in the form of single-event probabilities, for instance, 'The probability that this DNA match occurred by chance is 1 in 100,000'. That made it likely that jurors thought the defendant belonged behind bars. But when experts testified in frequencies instead, the case against the defendant appeared much weaker: 'Out of every 100,000 people, one will show a match'. Mathematically, that's the same, but, psychologically, it made jurors think about how many suspects there might be? In a city with one million adults, there should be 10 who match.⁹

CHAIR: I see. But what took defence teams so long to realize this?

HISTORIAN OF SCIENCE: There were a few attempts to introduce natural frequencies into court proceedings, even before 2000.¹⁰ Based on the laboratory research on natural frequencies, the O. J. Simpson defence team asked Judge Ito not to allow the prosecution's DNA expert, Professor Bruce Weir, to testify in terms of conditional probabilities and likelihood ratios, which are ratios of conditional probabilities. The defence requested frequencies instead. Judge Ito and the prosecution agreed, but the prosecution expert used likelihood ratios anyway! This was in the 'good' old days when statisticians like Weir didn't care about the psychology of jurors and judges.

⁹ Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review*, 67:859–886; Hoffrage, U., Lindsey, S., Hertwig, R. & Gigerenzer, G. (2000). Communicating statistical information. *Science*, 290:2261–2262.

¹⁰ See Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

SOCIOLOGIST: And this is when the French Feminist Association comes in.

STATISTICIAN: Because Simpson had beaten his wife?

SOCIOLOGIST: Ha, ha; very funny. No, it happened in 2010 at this very University. The Feminist Association had a larger goal . . .

HISTORIAN OF SCIENCE: Sorry, but the International Transparent Testimony Act was a few years earlier. After some local rulings by judges in England and the USA disallowing statements involving conditional probabilities or single-event probabilities in testimony, the International Federation of Law ruled that testimony in the courts had to be communicated in terms of natural frequencies rather than probabilities or likelihood ratios. That was in 2006, two decades after DNA fingerprinting was introduced into American criminal investigations. Probability statements about singular events are no longer admissible. The Act helped to bring insight into the court proceedings and get confusion out.

PSYCHOLOGIST: I am not a legal scholar, but isn't there, typically, more evidence than just a DNA match?

HISTORIAN OF SCIENCE: Yes, there is; but jurors and judges need to understand DNA evidence whether or not there is additional evidence, such as eye witness accounts. And since we have had complete DNA data banks in all European and North American countries for twenty years, there is a tendency for police officers to sit at their computers and search in data banks rather than at the scene of a crime.

SOCIOLOGIST: I am not persuaded that that law started the war against innumeracy. The Feminist Association was concerned for years with the harm done to women by misinformation about breast cancer screening. This harm included unnecessary anxiety and unnecessary surgery. For instance, in the year 2005, some 150,000 French women who did not have breast cancer were nevertheless operated on as a consequence of false positive mammogram diagnoses. These women were not informed that some 9 out of 10 positive screening mammograms are actually false positives, but believed that a positive test most likely meant that they had breast cancer. Even after it turned out that the positive result was a false positive, their anxiety remained. The Feminist Association discovered several other ways of presenting statistics with which the medical associations misled women. For instance the use of relative risks as opposed to absolute risks. Your mothers and grandmothers were told that mammography screening reduced mortality in women over 50 years of age by 25%. This is a relative risk. But 25% of what? Again, the reference class was not made transparent. Many women understood it to mean that, out of 1,000 women, 250 were saved from breast cancer. In fact, only 1 out of 1,000 was saved. That's the absolute risk: 0.1%.

CHAIR: How can 25% be the same as 0.1%?

PSYCHOLOGIST: Take 1,000 women who do not participate in screening. Four of them will die from breast cancer within 10 years. Now take 1,000 women who do participate. Three of them will die from breast cancer within 10 years. The difference between 4 and 3 amounts to a relative risk reduction of 25%. In absolute frequencies, however, that's 1 out of 1,000. Despite repeated calls from researchers that the women be given the information in absolute risks, they almost never were, they practically always received it in terms of relative risks. Just as the lack of information about false positives had unnecessarily increased women's anxiety, the information in terms of relative risk

reduction falsely increased women's faith in the benefits of screening. The Feminist Association was also concerned that the actual harm mammography screening can do was poorly communicated to women. They finally sued medical associations and industries around the world for intentionally producing confusion and innumeracy that violated the right of women to informed consent.

SOCIOLOGIST: This smart move on the part of the Feminist Association forced the World Health Organization into action. The national health associations and industries were sentenced by the courts to pay 10 billion Euros in damages.

STATISTICIAN: But it took several years before the WHO finally took action. There was a battle over whether the WHO should admit to, or even treat, innumeracy in physicians. Those in favour had all the evidence on their side: In study after study, 80 to 90% of physicians did not understand how to estimate the probability of a disease being present from, for example, standard screening tests.¹¹ Those in favour also had the therapeutic tools to cure innumeracy. The opponents did not want the medical profession to be associated with innumeracy in the public mind; they feared it would undermine its authority and the public's trust in it. The compromise was that innumeracy was declared a general mental aberration and it was entered into the DSM-VIII, thus it did not look like a doctor's disease alone, which it wasn't.

HISTORIAN OF SCIENCE: Yes, the WHO action was the turning point—that and the changes in the medical curricula in universities. The WHO did enter innumeracy into their catalogues of mental diseases and disabilities, but the definite change came with its Anti Mental Pollution Act, which made clear communication of risks the rule and eliminated forms of communication that had previously 'polluted' minds. Medical students learned how to express risks as absolute risks rather than relative risk reductions, and in natural frequencies rather than conditional probabilities; they learned how to specify reference classes, and so on. At the same time, high school curricula focused on training statistical thinking, and students learned how to play with representations. My own kids loved it when they could confuse people with percentages; this was much more fun than the applications of algebra and geometry.

CHAIR: Don't overlook the economic consequences of the Act. The Anti Mental Pollution Act and the fact that the WHO added innumeracy to their catalogue made millions of dollars available for research and implementation programmes. My estimate is that about 10 billion dollars alone were poured into education in professional schools and high schools.

HISTORIAN OF SCIENCE: But how do we know that all countries are innumeracy-free now? Can we trust these results?

STATISTICIAN: The Sixth International Mathematics and Science Study (SIMSS) focused exclusively on statistical thinking and defined innumeracy operationally in terms of performance on the test. The test items measured the ability to use representations to

¹¹ Hoffrage, U. & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, 73:538–540; Gigerenzer, G. (1996). The psychology of good judgment: Frequency formats and simple algorithms. *Journal of Medical Decision Making*, 16:273–280.

which the human mind is adapted, such as figures and natural frequencies. It also included test items that measured understanding uncertainties in the real world rather than hypothetical situations like urns and balls. SIMSS enabled each country's performance to be measured by the same standard, and, when 99% of twelfth graders passed the test, the country was declared free of innumeracy.

SOCIOLOGIST: What's with the rest of the population? The older people, those with chronic math anxiety, those who read poetry while others discuss baseball statistics?

STATISTICIAN: Worldwide, we reached the SIMSS performance level for the first time in 2054, this year. The professional schools in law, medicine, and business reached that level a decade ago.

SOCIOLOGIST: But still, what's with the older fellows, like you and me?

CHAIR: Time is up. And I still don't understand how 25% can be the same as 0.1%.

Note

By Professor Emile Ecu (Chair)

The editors of this book reminded me that psychology is a discipline that has a short-term rather than a long-term memory. That is, work older than 20 years—before 2034—is likely already out of memory. This is a most unfortunate state of affairs, which we do not have in economics. To compensate for this, I have added footnotes and references to clarify what the panelists refer to.

To begin with, the best analysis of the events 400 years ago that led to the mathematical theory of probability, and of the rise and decline of the classical interpretation of probability (in which frequencies and subjective degrees of belief were merged rather than distinguished), are the books by Lorraine Daston, *Classical probability in the enlightenment*, Princeton, NJ: Princeton University Press (1988) and Ian Hacking, *The emergence of probability*, Cambridge, England: Cambridge University Press (1975). Two hundred years ago, George Boole's *An investigation of the Laws of Thought on which are founded the mathematical theories of logic and probabilities* (New York: Dover 1854/1958) was published, a seminal work in which the laws of logic and probability are derived from the laws of psychology. One hundred years ago, Leonard J. Savage published *The foundations of statistics* (New York: Dover 1954), a seminal book that promoted a subjective view of probability and the revival of personal probabilities during the second half of the twentieth century.

CHAPTER 5

FREQUENCY JUDGEMENTS AND RETRIEVAL STRUCTURES: SPLITTING, ZOOMING, AND MERGING THE UNITS OF THE EMPIRICAL WORLD

KLAUS FIEDLER

Abstract

In this chapter, an attempt is made to explain the discrepancy between the often-cited finding of accurate and robust frequency memory and the existence of serious biases in frequentistic judgements. It is argued that accuracy often depends on the retrieval-cue structure. Accuracy is typically obtained when frequency judgements for different stimuli are elicited by comparable and easy-to-handle retrieval cues. However, strong biases arise when judgement prompts solicit different, misleading retrieval structures. Judges lack the metacognitive devices that would be necessary to correct for the influence of incomparable retrieval cues. The influence of retrieval structures on frequency judgements is illustrated with reference to various paradigms, such as category-split effects, the numerosity heuristic, conjunctions and logical connectives, the base rate fallacy, and inductive judgements of correlated attributes.

The assessment of frequencies in the empirical world is biologically significant and ubiquitous. Our acoustic system is equipped with sensors to assess the frequency of sound pressure waves; our emotional system is sensitive to the frequency of heartbeats and music beats per minute; and learning and conditioning are a function of the frequency of reinforcements. Even primitive organisms are able to learn the rate at which behaviours and situations are associated with pleasant (e.g. food) or unpleasant consequences (e.g. pain).

These elementary examples already illustrate a point that is easily overlooked: what is informative cannot be absolute, unconditional event frequencies per se. Rather, frequency information is only meaningful in the context of specific algorithms applied for frequency assessment. Thus, to perceive pitch, the absolute number of pressure changes must be normalized to some time unit and acoustic waves must co-occur in space. That is, an algorithm is required that assesses the number of pressure changes conditional on