Subjective Probability

Edited by

George Wright

University of Strathclyde in Glasgow, UK

and

Peter Ayton

City University, London, UK

Why the Distinction between Single-event Probabilities and Frequencies is Important for Psychology (and Vice Versa)

Gerd Gigerenzer
University of Chicago

Some years ago in Stanford I was lunching with a motley group of colleagues, mostly psychologists and economists, all interested in judgment under uncertainty. We gnawed our way through our sandwiches and through the latest embellishments of the prisoners' dilemma, trading stories of this or that paradox or stubborn irrationality. Finally, one economist from Princeton concluded the discussion with the following dictum: "Look," he said with conviction, "either reasoning is rational *or* it's psychological."

This forked opposition between the rational and the psychological has haunted me ever since. Frege scholars will hear in it an echo of the nineteenth-century debate between the logician Frege and the psychologist Wundt over the status of the "laws of thought"; the economists and psychologists seated at the picnic table with me that afternoon had in mind the more recent findings of the "heuristics and biases" research program in cognitive psychology (e.g. Tversky & Kahneman 1974, 1983). Certainly anyone acquainted with only this aspect of contemporary psychology—and it remains among the best publicized, both to colleagues in other disciplines and to the public at large—could

easily have come to think that psychology is about revealing and explaining human irrationality. The *conjunction fallacy*, the *base-rate fallacy*, the *overconfidence bias*—this was the gloomy litany of sins people seemed to commit routinely and incorrigibly against reason. According to the exponents of the "heuristics and biases" program, human beings were programmed to be systematically, stubbornly *irrational* when making judgments under uncertainty—at least, most of the time. (Experimental subjects were not dazzling at logical thinking either, but that is another story and another research program.) No wonder the psychology of reasoning had become nearly synonymous with the investigation of the irrational (you get a taste from Table 7.1).

What exactly did it mean to be irrational, according to the psychologists of the heuristics and biases program? Let me use a well-known example, the "Linda problem". Assume you are a subject in a psychological experiment. In front of you is a text problem and you begin to read:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination

Table 7.1 A sample of conclusions from the *heuristics and biases* program

In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors.
<div align="right">Daniel Kahneman & Amos Tversky (1973, page 237)</div>

It appears that people lack the correct programs for many important judgmental tasks. ...we have not had the opportunity to evolve an intellect capable of dealing conceptually with uncertainty.
<div align="right">Paul Slovic, Baruch Fischhoff & Sarah Lichtenstein (1976, page 174)</div>

The genuineness, the robustness, and the generality of the base-rate fallacy are matters of established fact.
<div align="right">Maya Bar-Hillel (1980, page 215)</div>

The biases of framing and overconfidence just presented suggest that individuals are generally affected by systematic deviations from rationality.
<div align="right">Max Bazerman & M.A. Neale (1986, page 317)</div>

[Overconfidence bias] has proved so robust that it is hard to acquire much insight into the psychological processes producing it.
<div align="right">Baruch Fischhoff (1988, page 172)</div>

[We are] a species that is uniformly probability-blind, from the humble janitor to the Surgeon General ... We should not wait until A. Tversky and D. Kahneman receive a Nobel prize for economics. Our self-deliberation from cognitive illusions ought to start even sooner.
<div align="right">Massimo Piattelli-Palmarini (1991, page 35)</div>

and social justice, and also participated in antinuclear demonstrations. Which of these two alternatives is more probable?

(a) Linda is a bank teller
(b) Linda is a bank teller and active in the feminist movement.

Which alternative would you choose? Assume you chose (b), just as most subjects—80% to 90%—in previous experiments did. Tversky and Kahneman (1983) argue: (b) is the conjunction of two facts, namely that Linda is a bank teller and is active in the feminist movement, whereas (a) is one of the conjuncts. Because the probability of a conjunction cannot be greater than that of one of its conjuncts, the correct answer is (a), not (b). Therefore, your judgment is recorded as an instance of a celebrated reasoning error, known as the *conjunction fallacy*. Tversky, Kahneman, and others have shown that this type of judgment is highly stable across experimental manipulations. By analogy to stable visual illusions, stable reasoning errors, such as the conjunction fallacy, have been labeled *cognitive illusions*. The standard conclusion is that the mind does not possess the proper statistical algorithms, but relies on non-statistical quick-and-dirty algorithms, such as the *representativeness heuristic*. That is, the mind assesses the probability by calculating the similarity between the description of Linda and each of the alternatives, and chooses the alternative with the highest similarity. Judging probability by similarity has been termed the *representativeness heuristic*.

This alleged demonstration of human irrationality in the Linda Problem has been widely publicized in psychology, philosophy, economics and beyond. Stephen J. Gould (1992, page 469) puts the message clearly:

> I am particularly fond of [the Linda] example, because I know that the [conjunction] is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me—"but she can't just be a bank teller; read the description." ... Why do we consistently make this simple logical error? Tversky and Kahneman argue, correctly I think, that our minds are not built (for whatever reason) to work by the rules of probability.

In what follows I will argue that Gould should have had more trust in the intuition of his homunculus.

Two aspects of the standards of rationality versus irrationality assumed by this and other celebrated experiments cry out for closer inspection.

(1) *The distinction between single-event probabilities and frequencies.* In the supposed demonstrations of the conjunction fallacy, the base rate fallacy, and the overconfidence bias, rationality is characterized not simply by probability theory, but some particular *interpretation* of probability theory, often a narrow version of Bayesianism. In the Linda problem, probability theory

is applied to a single event—that Linda is a bank teller—rather than to frequencies. Does probability theory apply to single events?

This is a controversial matter amongst probabilists, who have long and heatedly debated the merits of subjective Bayesian versus objective frequentist interpretations of probability. The influential Bayesian Leonard J. Savage (1954), for instance, introduced his notion of personal probability with everyday examples of reasoning about *singular* events: "I personally consider 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. But even this late spring snow seems to me more probable than that Adolf Hitler is still alive" (page 27). Savage's proposal challenged the frequentist schools which were then dominant, as they are in most statistics departments today. Savage was quite explicit about the deviant character of his proposal, when he added, "Many, after careful consideration, are convinced that such statements about probability to a person mean precisely nothing, or at any rate that they mean nothing precisely" (page 27).

The mathematician Richard von Mises (1957) was one of those many. In his view, a reference class (collective) has to be defined first, and then the probability of a *repetitive* event is the relative frequency of this event in its class. One of his examples is the probability of death at age 40, as determined from the data of insurance companies. The class is "all men insured before reaching the age of forty after complete medical examination and with the normal premium". The number of deaths at age 40 was 940 out of 85 020, which corresponds to a relative frequency of about 0.011. This probability is attached to a class, but not to a particular person or a single event. Every particular person is always a member of many different classes, whose relative frequencies of death may have different values. Therefore, von Mises concluded, "It is utter nonsense to say, for instance, that Mr. X, now aged forty, has the probability 0.011 of dying in the course of the next year." (pages 17–18).

By now it should be clear that according to a strong frequency view of probability (e.g. Neyman, 1977; von Mises, 1957), what has been labeled the conjunction fallacy is *not* an error in probabilistic reasoning. In this view, probability theory is about frequencies and simply doesn't apply to single events.

(2) *Content-independent rationality.* There is a peculiar indifference in this standard of rationality to background knowledge: rationality here means the deployment of formal algorithms (or rules, such as the conjunction rule) which are content-independent. That is, it is assumed that they can and should be applied to tasks with different specific contents, provided the formal structure remains constant. From this point of view, rationality is not bound to any specific domain, and knowledge ideally is irrelevant to proper reasoning.

In the Linda problem, for instance, whatever you know about bank tellers and feminists is assumed to be entirely irrelevant; indeed, you need not read

the description of Linda at all—it is irrelevant to content-independent rationality. Hence there is little analysis of how the content of a problem cues the understanding of the term "probable" in this research tradition. "Probable" can refer to *typical, prototypical, frequent, credible,* to the *weight of evidence,* to a *plausible causal story,* or to what "may in view of present evidence be reasonably expected to happen," as the *Oxford English Dictionary* informs us. Most of these uses do not obey the laws of probability. For instance, judgments of typicality do *not* follow the conjunction rule. Betty Friedan may count as a typical *feminist writer,* but not as a typical *writer.* Such psychological considerations, however, are not part of the content-independent rationality that defines right and wrong reasoning in the heuristics and biases program.

These two issues are not independent. For instance, suppose one insists that every single-event statement involving the term "probable," as in the Linda problem, must obey the laws of probability theory rather than, say, the guidelines of the *Oxford English Dictionary.* (I take the statements in Table 7.1 to epitomize this conviction.) One would then be uninterested in how content (add physical and social context, goals, if you want) determines what is reasonable in a given situation.

I will focus in this chapter on the distinction between single-event probabilities and frequencies, and will say little about the role of content in understanding what is rational (on the latter see Cosmides & Tooby, 1992; Gigerenzer, 1991; Gigerenzer & Hug, 1992).

In the first part, drawing on recent work in the history of probability, I will show that the distinction between single-event probabilities and frequencies was dependent on theories of mind: the meaning of probability changed when theories of mind changed. In the second part, drawing on recent experimental work, I will show that apparently stable cognitive illusions are dependent on the distinction between single-event probabilities and frequencies: cognitive illusions tend to disappear when single-event probabilities·are changed into frequencies. Thus, I argue that the conceptual distinction between single-event probabilities and frequencies is of direct relevance for psychology, and *vice versa.*

7.1 HOW THEORIES OF PSYCHOLOGY SHAPED THE MEANING OF PROBABILITY

According to legend, probability is one of the few seminal ideas that has an exact birthday. In 1654, precisely three hundred years before Savage's treatise, the now famous correspondence between Blaise Pascal and Pierre Fermat first cast the calculus of probability in mathematical form. Ian Hacking (1975) argued that the notion of probability that emerged so suddenly was

Janus-faced from the very beginning. One face was aleatory, concerned with observed frequencies (e.g. co-occurrences between fever and disease, comets and deaths of kings); the other face was epistemic, concerned with degrees of belief or opinion warranted by authority. In his view, the twentieth-century duality between objective frequencies and subjective probabilities existed then as it does now. Barbara Shapiro (1983) and Lorraine Daston (1988), however, have argued that probability in the seventeenth and eighteenth centuries had more than Janus's two faces. It included physical symmetry (e.g. the physical construction of dice, now called "propensity"); frequency (e.g. how many people of a given age died annually); strength of argument (e.g. evidence for or against a judicial verdict); intensity of belief (e.g. the firmness of a judge's conviction in the guilt of the accused); verisimilitude; and epistemological modesty, among others. Over the centuries, probability also conquered new territory and created further meanings, such as in quantum physics, and lost old territory, such as the probability of causes (Daston, 1988). Rather than Janus's two faces, probability seems more like a group of visages loosely assembled in a family portrait, with some members joining over time and others dropping out.

7.1.1 The Unity: Frequencies *and* Subjective Beliefs

The puzzling fact about the Enlightenment probabilists is *the ease with which they slid from one meaning of probability to the next*—and this holds independently of whether you see probability as Janus-faced or more like a family portrait. This ease created the apparent paradox that competing present-day interpretations of probability could claim the same work as their ancestor. Jakob Bernoulli's *Ars conjectandi* (1713), for instance, has been variously claimed as anticipating the 20th century's subjective interpretation, Rudolf Carnap's logical interpretation, and the extreme frequentist inter-pretation of Jerzy Neyman and Richard von Mises (Hacking, 1975, pages 15−16).

   The solution to this puzzle lies in the intimate link between psychology and probability. Daston (1988, Chapter 4) argues that only with hindsight does it seem that Bernoulli and other classical probabilists vacillate between objective and subjective interpretations. Whereas today these interpretations look incompatible to many, the classical probabilists were able to reconcile the subjective and objective facets of probability on the basis of the theories of mind advanced by John Locke, David Hartley, and David Hume. The following account of how associationist psychology shaped and incorporated ideas of probability is a condensed version of Daston's (1988) detailed study.

   Philosophers such as Hartley and Hume, and mathematicians like Condorcet and Laplace, treated associationist psychology and mathematical probability as kindred topics. Following Locke's associationism, Hume held that the mind unconsciously and automatically tallied frequencies and

proportionated degree of belief (for Hume, the *vivacity* of an idea). And Hume insisted that the psychological mechanism that converted frequency into belief was finely tuned: "When the chances or experiments on one side amount to ten thousand, and on the other to ten thousand and one, the judgment gives the preference to the latter, upon account of that superiority." (Hume, 1739/1975, page 141). Despite his reservation about the validity of induction, Hume made probabilistic thinking the *de facto* standard of reasonableness. Hume linked frequency with belief, but his account contained almost no reference to the mathematical theory of probability. David Hartley's (1749) work did. He combined elements from Locke's sketch of associationism and Newton's physiological speculations concerning the vibratory basis of sensations, and worked it into a full-blown associationism that connected the laws of mind with the laws of probability. Repeated associations created cerebral vibrations until grooves of mental habit were etched in the brain. Through this physiological mechanism, human judgment, when undeflected by strong emotion or passion, imitated the law of large numbers.

The list of psychological mechanisms underlying the mapping of objective frequencies into subjective belief, postulated from Locke to Hartley to Laplace, seems surprisingly familiar to a contemporary psychologist: observed frequencies are transformed into degrees of belief through "traces", "vibrations", "interior images", and "impressions". All these mechanisms assumed the passive, automatic and unconscious mapping of experienced frequencies into subjective probabilities. Being built up from frequencies, degrees of belief were considered to be rational. The Enlightenment empiricists had taken due notice of the distortion of rational belief through passion and interest, but they believed these were corrigible aberrations.

These psychological theories were the backbone of what is now known as the *classical interpretation of probability* (from 1660 to c. 1840) and they explain some of its central features. First, classical probability conflated subjective belief and objective frequencies, based on associationist psychology. Second, probabilities were epistemic, a figment of human ignorance and therefore subjective, not part of the physical world. Classical probabilists, from Jakob Bernoulli through Laplace, were arch determinists (Daston, 1992). God, or Laplace's secularized demon, could dispense fully with probability. However, we humans are, as John Locke put it, most of the time condemned to live in the twilight of probability rather than the noonday sun of certainty. Although the world itself is deterministic, human cognition is inherently probabilistic and empirical in its working—a view that was revived, among others, in Egon Brunswik's (1955) functional probabilism. Third, the mapping of frequencies into subjective probabilities was considered to be rational, as were subjective probabilities, unless disrupted by passion or interest. The Enlightenment probabilists cherished the fiction of the *hommes éclairés*, an elite of educated people who could prevent such disruptions from affecting their beliefs. Probability theory mirrored their reasoning and provided a tool

for those unfortunates needing help to stay clear of these disruptions. Human reasoning and probability theory were two sides of the same coin. In Laplace's famous phrase, probability theory was nothing more than "good sense reduced to a calculus".

By the time Siméon-Denis Poisson (1837) published his major work on probability, the classical interpretation was under attack on several fronts. The psychological theories postulating mechanisms that guaranteed the proportioning of belief to frequencies had given way to those that emphasized the illusionary nature of human belief. Etienne de Condillac (1754) was one of the first to express misgivings about the reliability of the link between frequency and belief. In his psychology, wishful thinking became the rule rather than the exception. Condillac was preoccupied with pathological associations caused by experiences early in life, by prejudice, or by brain consistency. He held, for instance, that young girls were prone to confuse chimeras for realities, because their brains were soft and even faint associations left permanent impressions in a soft medium. Condillac and his followers shifted the associationist psychology of Hume and Hartley to a psychology in which needs, wants, and temperaments (and other sources of pathologies) determined how the mind distributed attention, which in turn organized experience (Daston, 1988). The unity between frequency and belief was slowly eroded. What psychology had given to probability, it now took away. Poisson was the first to distinguish clearly in print, in 1837, between the *subjective* and the *objective* meaning of probability.

There is a broader intellectual and social context in which the rise and the fall of the classical interpretation of probability is embedded. The French revolution and its aftermath seems to have shaken the confidence of the probabilists in the existence of a single shared standard of reasonableness. What constituted "good sense" was no longer self-evident. The consensus and the values of the intellectual and political élites fragmented and disappeared, as did *l'homme éclairé*, the fiction of the reasonable man who embodied this consensus (Daston, 1988; Gigerenzer *et al.*, 1989, Chapter 1).

7.1.2 The Divorce: Frequencies *versus* Subjective Belief

Subjective belief and objective frequencies began as equivalents and ended up as diametric opposites. Poisson had distinguished the two, and the political economist and philosopher Antoine Cournot (1843) seems to have been the first who went one step further and eliminated subjective belief from the realm of mathematical probability: mathematical probability was not a measure of belief. Only then did it become evident that the classical interpretation of probability had been an *interpretation*. Classical probability was a form of "mixed mathematics", a term stemming from Aristotle's

explanation of how optics and harmonics mixed the forms of mathematics with the matter of light and sound. Classical probability theory had no existence independent of its subject matter—the beliefs of reasonable men. The modern view that a mathematical theory might indeed exist independently of a particular subject matter—the distinction between formal theory and application—was foreign to mixed mathematics. Arguably, mathematical probability did not free itself from its particular applications until very recently, when in 1933 A.N. Kolmogoroff presented his axiomatization of probability.

The new associationist psychology which focused on illusions had, by the early nineteenth century, provided the arguments for severing subjective probabilities from objective frequencies, and, ironically, associationist psychology from probability theory. By about 1840, *l'homme éclairé* had given way to *l'homme moyen*. Probability was no longer about mechanical rules of rational belief embodied in an élite of reasonable men, but about the properties of the average man (*l'homme moyen*), the embodiment of mass society if not mediocrity. Adolph Quetelet's (1835) *social physics* determined the statistical distributions of suicide, murder, marriage, prostitution, height, weight, education, and almost everything else in Paris, and compared these with the distributions in London or Brussels. The means of these distributions defined the fictional *average man* in each society. The means and rates of moral behaviors, such as suicides and crimes in Paris or in London, proved to be strikingly stable over the years; this was cited as evidence that moral phenomena are governed by the laws of a society rather than by the free decisions of its individuals. In nineteenth-century France, statistics became known as "moral science".

This new focus on mass phenomena had a tremendous impact on pioneer sociologists such as Herbert Spencer and Emile Durkheim, and shaped demography, insurance, epidemiology, Prussian bureaucracy, the debates on free will, Francis Galton's enthusiasm for the normal curve and Gustav Theodor Fechner's statistical aesthetics, *inter alia* (Hacking, 1990; Stigler, 1986). Quetelet's model of human behavior as erratic and unpredictable at the individual level, but governed by statistical laws and predictable at the level of society, was independently adopted by James Clerk Maxwell and Ludwig Boltzmann to justify, by analogy, their statistical interpretation of the behavior of gas molecules (Porter, 1986). By this strange route, physics became revolutionized through the analogy with statistical laws of society.

Throughout most of the nineteenth and twentieth centuries, the "probabilistic revolution" (Krüger, Daston & Heidelberger, 1987; Krüger, Gigerenzer & Morgan, 1987) was about frequencies: from the kinetic theory of gas to quantum statistics, and from population genetics to the Neyman–Pearson theory of hypothesis testing. The urn model of classical probability was now concerned with these mass phenomena, excluding subjective degrees of belief

such as single-event probabilities. Joseph Bertrand in his *Calcul des probabilités* (1889), for instance, criticized Laplace's applications of Bayes' theorem to calculate degrees of belief: we believe the sun will rise tomorrow because of "the discovery of astronomical laws and not by renewed success in the game of chance" (page xliv).

As is well known, subjective probability has regained acceptance in the second half of this century with the pioneering work of Bruno de Finetti and Frank Ramsay in the 1920s and 1930s and Leonard Savage in the 1950s. The reasonable man, once exiled from probability theory, had his comeback. Economists, psychologists, and philosophers now struggle again with the issue of how to codify "reasonableness" in mathematical form—the same issue once abandoned by mathematicians as a thankless task. Before the 1970s, the return of subjective probability still provoked a particularly lively debate between frequentists and subjectivists (now called "Bayesians"). Today, both sides pretend to know each other's arguments all too well and seem to have stopped listening. Frequentists dominate statistics and the experimental sciences; subjectivists the areas of theoretical economics and artificial intelligence. The territory has been divided up. As Glenn Shafer (1989) complained, "conceptually and institutionally, probability has been balkanized" (page 15).

To summarize: theories of psychology have been important in shaping the meaning of probability, and therewith the subject matter of probability theory. In particular, the associationist psychology of Locke, Hume and Hartley provided the grounds for *not* distinguishing objective frequencies and subjective degrees of belief—from the inception of probability theory *circa* 1650 to roughly 1840. The turn of associationist psychology towards illusions dethroned the reasonable man of classical probability theory and made the distinction between degrees of reasonable belief and frequencies obvious. After this conceptual transformation, psychology found itself dissociated from probability theory, too.

7.2 HOW THE DISTINCTION BETWEEN SINGLE EVENTS AND FREQUENCIES AFFECTS COGNITIVE ILLUSIONS

Psychologists like precise birthdays too. Textbooks celebrate 1879 as the beginning of what is referred to as *scientific psychology*, when Wilhelm Wundt devoted some space at the University of Leipzig for conducting experiments. For Wundt, the experimental method was a means to study elementary cognitive processes, such as attention and perceptual thresholds, but not (what he believed to be) deeply culture-bound processes such as thinking (Danziger, 1990). For these and other reasons, such as the dominance of American

behaviorism, probabilistic reasoning was only occasionally a topic for psychologists in the first half of this century.

The classical probabilists would have felt a strong sense of *déjà vu* upon learning about some of the theoretical developments in the second half of the twentieth century. Around 1950, Jean Piaget in Geneva revived the reasonable man of classical probability theory. In Piaget and Inhelder's (1951/1975) experimental work, the formal laws of probability are the laws of the adolescent and adult mind. Errors in probabilistic reasoning were characteristic only during ontogenetic development, until the age of fourteen or so, when formal probabilistic reasoning emerges. Take for instance the law of large numbers. In 1703 Jacob Bernoulli had written in a letter to Leibniz that the law of large numbers is a rule that "even the stupidest man knows by some instinct of nature *per se* and by no previous instruction" (see Gigerenzer et al., 1989, page 29). More than two centuries later, Piaget and Inhelder concluded that even twelve to thirteen year olds intuitively apply the law of large numbers and understand the reasons for the law (page 207).

Locke, Hartley, and Hume had assumed that the mind unconsciously tallies frequencies and converts them into degrees of belief. Hasher and Zacks (1979) concluded from their experiments that frequencies are one of the few kinds of information (the others being word meaning and spatial and temporal location) that are monitored *automatically*—that is, without intention or much attention, and without interfering with other tasks. Moreover, what is now called *automatic frequency processing* seems to be generally accurate, a conclusion independently arrived at by others (e.g. Brehmer & Joyce, 1988). The thesis that objective frequencies eventually shape degrees of belief has now been experimentally demonstrated (Hasher, Goldstein & Toppino, 1977). Locke, Hartley, and Hume would have been enthusiastic about these experimental findings. The reasonable man is back, dressed in modern fashion: less élite (everyone is a reasonable intuitive statistician) and confirmed by numerous experimental results.

The *déjà vu*, however, goes beyond the recreation of the reasonable man. Around 1970, much of cognitive and social psychology turned away from the rational intuitive statistician and focused on illusions (Kahneman, Slovic & Tversky, 1982; Nisbett & Ross, 1980). One and a half centuries earlier, associationist psychology had turned to illusions, and the reasonable man had crumbled along with the classical interpretation of probability. Now illusions were again being used to destroy belief in the rational *homo sapiens*, and to challenge economists' rational *homo economicus*. Now, as then, illusions were no longer the exception, but the rule.

Here the historical parallels end. The old challenge was that passion and wishful thinking almost always interfere with the rational laws of thought. Freud's attack on human rationality is a well-known variation on that theme. The unconscious wishes and desires of the Id are a steady source of

intrapsychic conflict that manifests itself in all kinds of irrational beliefs, fears, and behaviors. The new challenge, however, does not invoke passion or wishful thinking as interfering with otherwise rational reasoning. This challenge is stronger: the human mind does not possess the proper statistical algorithms. Poor reasoning is seen as a straightforward consequence of the laws of human reasoning, which are non-statistical "rule-of-thumb" heuristics. The mind is a poor intuitive statistician whether or not passion and wishful thinking compound this state of affairs.

Ironically, the departure point of the *unreasonable man* that emerged two decades after Piaget's revival of the *reasonable man* was Savage's neo-Bayesianism. In the 1960s, Ward Edwards and his colleagues at the University of Michigan made two related proposals. First, Edwards, Lindman & Savage (1963) attempted to persuade experimental psychologists to turn Bayesian and to dispense with frequentist hypothesis testing. Second, Edwards (1968) proposed to study empirically whether intuitive reasoning follows Bayesian statistics. The first proposal fell still-born from the press; the second became a raging success.

Experimenters already had their frequentist statistics, a curious and con-fused mishmash of Fisher's significance testing and Neyman–Pearson's hypotheses testing (Gigerenzer, 1993). This was generally presented as the *sine qua non* of scientific method. Textbooks did not tell their readers that they were teaching a shotgun marriage between Fisher and Neyman–Pearson. Rather, the textbooks created the illusion that "statistics is statistics is statis-tics". Since the 1950s, statistical inference had become a mechanical ritual in psychology and beyond, enforced by journal editors and internalized by researchers as the guardian of objectivity and scholarly morality. Bayesianism, by contrast, looked subjective and, above all, unnecessary.

Thus, in the 1970s and 1980s, Bayesianism became a rational yardstick for the *subjects* in psychological experiments, but not for the *experimenters* who analyzed them. Subjects were judged rational if their inferences from data to hypotheses followed Bayes' theorem; otherwise their judgments were recorded as errors in reasoning, such as the *base rate fallacy* (see below). However, when experimenters made inferences from data to hypotheses— here, whether subjects are Bayesians—they did *not* use Bayes' theorem. They used, as they had been taught for two decades before Edwards' proposal, frequentist statistics. But the most commonly used kind of frequentist statistics, R.A. Fisher's significance testing, does not use prior probabilities or base rates. This neglect of base rates by experimenters was not recorded as an error in reasoning, although it had all the characteristics of the base rate fal-lacy. Nor do I know of a single experimenter who noticed and remarked on that amazing double standard. The split between Bayesians and frequentists not only divides disciplines today, but can also go right through a single individual (Gigerenzer, 1993).

Edwards seems to have soon become dissatisfied with pointing out discrepancies between subjects' reasoning and Bayes' formula, and no interesting and rich theory of how subjects actually do reason had emerged. He turned to the task of designing tools that help people reason the Bayesian way. In the 1970s, Amos Tversky and Daniel Kahneman took over Edwards' second proposal and turned it into what is now known as the *heuristics and biases* program.

The heuristics and biases program arrived at a view of human rationality (Table 7.1) diametrically opposed to that of classical probability theory. Yet this modern program neglects the distinction between single-event probabilities and frequencies just as the classical probabilists ignored the distinction between subjective degrees of certainty and objective chances.

I will now use the distinction between single-event probabilities and frequencies to unearth the reasonableness hidden by the perspective of the heuristics and biases program.

7.2.1 Representation of Information: Single-event Probabilities versus Frequencies

My point here is precisely *not* to champion one side over another—frequentism over Bayesianism, or vice versa—but to point out a connection between the single-event probabilities/frequencies distinction and a second distinction, that between algorithms and information representation.

Much ink flowed in debates about mental *algorithms*: is the mind equipped with the right statistical algorithms or only with suboptimal algorithms based on rules of thumb such as the representativeness heuristic? These two alternatives, however, are not sufficient for a theory of cognitive processes underlying judgment under uncertainty, because they only deal with the level of algorithms. Algorithms need *information*, and information needs *representation*. This distinction between algorithms and information representation is central to David Marr's (1982) analysis of visual information-processing systems.

For example, consider *numerical* information. This information can be represented by the Arabic numeral system, the binary numeral system, Roman numerals, and other symbol systems. These different representations can be mapped *one-to-one onto each other*, and are in this sense formally equivalent representations. But they are not necessarily equivalent for calculating algorithms. The algorithms programmed into my pocket calculator work well when I feed them Arabic numerals, but not when I feed them binary numbers. The human mind seems to acquire analogous preferences for one form of representation: contemplate for a moment long division in Roman numerals.

Let me now return to the distinction between single-event probabilities and frequencies. Instead of squabbling over which captures the "real" meaning of probability, let us instead regard them as two different representations of probability information. Finer distinctions can be made, but this will suffice to start.

An evolutionary speculation links these two distinctions. Assume that some capacity or algorithm for statistical reasoning has been built up through evolution by natural selection. For what information representation would such an algorithm be designed? Certainly not for percentages and single-event probabilities (as is assumed in many experiments on human reasoning), since these took millennia of literacy and numeracy to evolve as tools for communication. Rather, in an illiterate and innumerate world, the representation would be frequencies of events, sequentially encoded as experienced—for example, *3 out of 20* as opposed to 15% or $p = 0.15$. Such a representation is couched in terms of discrete cases, that is, natural numbers.

Note that bumblebees, birds, rats, and ants all seem to be good intuitive statisticians, highly sensitive to changes in frequency distributions in their environments, as recent research in foraging behavior indicates (Gallistel, 1990; Real & Caraco, 1986). One wonders, reading that literature, why birds and bees seem to do so much better than humans.

In short, the proper functioning of a mental algorithm depends on the way in which information is represented. So, to analyze probabilistic reasoning, we must attend to the difference between, at least, the frequency and the single-event representation of probability. If evolution has favored one of these forms of representation, then it would be frequencies, which prelinguistic organisms could observe and act on.

Attending to this distinction suffices to make several apparently stable cognitive illusions disappear.

7.2.2 How to Make the Conjunction Fallacy Disappear

Now we apply the distinction between single-event and frequency information representation to the Linda problem. We only change the format from single event to a frequency representation, leaving everything else as it was.

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.
   There are 100 people who fit the description above. How many of them are:
   (a) bank tellers
   (b) bank tellers and active in the feminist movement.

Subjects are now asked for frequency judgments rather than for single-event probabilities. If the mind solves the Linda problem by using a representativeness

heuristic, changes in information representation should not matter, because they do not change the degree of similarity. The description of Linda is still more representative of (or similar to) the conjunction "teller and feminist" than of "teller." Subjects therefore should still commit the conjunction fallacy.

However, if there is some statistical algorithm in the mind *that is adapted to frequencies* as information representation, then something striking should happen to this stable cognitive illusion. Violations of the conjunction rule should largely disappear.

Table 7.2 How to make the *conjunction fallacy* disappear

Linda problem	Conjunction violations (%)
Single-event versions	
Tversky & Kahneman (1983)	
Which is more probable?	85
Probability ratings	82
Probability ratings T^*	57
Betting	56
Fiedler(1988)	
Probability ranking, Exp. 1	91
Probability ranking, Exp. 2	83
Hertwig & Gigerenzer (1993)	
Probability ranking	88
Frequency versions	
Fiedler (1988)	
How many out of 100?	22
How many out of X?	17
Hertwig & Gigerenzer (1994)	
How many out of 200?	13
How many	16

Note: The various versions of the Linda problem are (i) which is more probable (see text, $n = 142$), (ii) probability ratings on a nine-point scale ($n = 119$), (iii) probability ratings using the alternative "Linda is a bank teller whether or not she is active in the feminist movement" (T^*) instead of "Linda is a bank teller" (T) ($n = 75$), (iv) hypothetical betting, i.e. subjects were asked "if you could win \$10 by betting on an event, which of the following would you choose to bet on?" ($n = 60$). Fiedler asked subjects to rank order T, T&F, and other alternatives with respect to their probability. In his first frequency version the population size was always 100, in the second it varied ($n = 44$ and 23, in Experiments 1 and 2, respectively). Hertwig & Gigerenzer asked subjects to rank order T, T&F, and F, with respect to their probability (single-event version, $n = 24$), or estimate the frequency of T, T&F, and F (in the two frequency versions, each $n = 25$). In one of the frequency versions, the number of women was specified (200); in the other, this number was not specified.

The available experimental evidence confirms this prediction. Klaus Fiedler (1988) reported that the number of conjunction violations in the Linda problem dropped from 91% in the original, single-event representation to 22% in the frequency representation. A similar result was found when he replaced "there are 100 people" by some odd number such as "there are 168 people." The drop in the number of conjunction violations here was from 83% to 17%. Hertwig and Gigerenzer (1994) used three alternatives: F (Linda is active in the feminist movement), T&F (Linda is a bank teller and active in the feminist movement) and T (Linda is a bank teller). In the single-event task, subjects rank-ordered F, T&F and T with respect to their probability; in the frequency task, they estimated the frequency of T, T&F and F ("how many out of 200?"). The percentage of conjunction violations dropped from 88% in the single-event task to 13% and 16%, respectively, in two frequency tasks.

Hertwig and Gigerenzer as well as Fiedler reported similar results for other reasoning tasks from which the conjunction fallacy has been inferred as a stable cognitive illusion. Tversky and Kahneman (1983) had reported a similar case in their original paper, but maintained the claim that people commit a fallacy when choosing the conjunction in the single-event case.

To summarize: The philosophical and statistical distinction between single events and frequencies clarifies that judgments hitherto labeled instances of the "conjunction fallacy"' cannot be properly called reasoning errors in the sense of violations of the laws of probability. The conceptual distinction between single-event and frequency *representations* suffices to make this allegedly stable cognitive illusion largely disappear. The conjunction fallacy is not the only cognitive illusion that is subject to this argument.

7.2.3 How to Make the Base-rate Fallacy Disappear

In the 1960s, Ward Edwards and his colleagues designed probability revision problems to find out whether their subjects were Bayesians. Many of these problems used the tried-and-true urns-and-balls problems, and the major finding was that subjects exhibited *conservatism*—that is, that they seemed to give too much weight to the base rates. From the 1970s on, however, Tversky, Kahneman, and many of their followers claimed that reasoning deviates from Bayes' rule in the opposite direction, that subjects in fact ignore base rates—the so-called *base-rate fallacy*.

Recently, some researchers have weakened their claims about the generality and robustness of the base-rate fallacy, but some of the fundamental confusions with which this stimulating research was burdened from the very start have survived (Gigerenzer & Murray, 1987, ch.5).

The two confusions I will point out are both instances of blurring single-event probabilities and frequencies. The first confusion is between the Bayesian notion of a person's *prior probability* and the frequentist concept of

a *base rate*. Tversky and Kahneman (e.g. 1974; Kahneman & Tversky, 1973) started out using the expressions "neglect of base rates" or "insensitivity to base rates" interchangeably with those of "neglect of prior probabilities" or "insensitivity to prior probabilities." However, priors and base rates are different things. Priors are subjective degrees of belief that may be informed by objective base rates, but need not be identical. (Similarly, the subjective likelihoods that enter Bayes' theorem and the "individuating" information presented by the experimenter need not be identical; see Birnbaum, 1983; Schum, 1990). This confusion, however, was necessary to argue that if a subject does not give much weight to whatever base rate information the experimenter has presented, this counts as a demonstration of a fallacy, i.e., that the subject does not reason by Bayesian principles. Whether or not the mind actually reasons by Bayesian principles, this confusion between a *base rate* and a *subjective prior* has prevented us from drawing adequate conclusions from experimental work.

The second and related confusion is between normative theories of the subjective and frequentist varieties. For instance, when subjects seemed not to pay much attention to base rate information, Kahneman and Tversky (1973, p. 243) asserted: "The failure to appreciate the relevance of prior probability in the presence of specific evidence is perhaps one of the most significant departures of intuition from the normative theory of prediction." But *which* normative theory? They seem to have had Bayesianism in mind, and at that, a narrow version thereof—e.g., one that conflated base rates with priors (Gigerenzer, Hell & Blank, 1988). But what if intuition were measured against the frequency view?

I will now apply the distinction between single-event and frequency information representation to the base-rate fallacy . Here is an observation to start with.

Some researchers tend to *change* the representation of a problem from single-event probabilities to frequencies when they turn away from their subjects and explain the correct solution to their readers. An early example is Hammerton (1973, p. 252) who used *single-event probabilities* to communicate information to his subjects:

1. A device has been invented for screening a population for a disease known as psylicrapitis. 2. The device is a very good one, but not perfect. 3. If someone is a sufferer, there is a 90% chance that he will be recorded positively. 4. If he is *not* a sufferer, there is still a 1% chance that he will be recorded positively. 5. Roughly 1% of the population has the disease. 6. Mr. Smith has been tested, and the result is positive. The chance that he is in fact a sufferer is:

Hammerton seems to have been surprised that his subjects gave a median response of 85% (which is close to the 90% hit rate) despite the 1% base rate.

Such judgments have been labeled by others as the base-rate fallacy. When the author explained the correct answer to his readers, he switched however, without comment into a *frequency* representation (p. 252):

> Out of every 100 persons tested, we expect 1 to have the disease; and the device is nearly certain to say that he has. Also, out of that 100, we expect the machine to say that 1 healthy person has the disease. Thus, in the long run, out of every 100 persons tested, we expect 2 positive results, one of which will be correct and the other incorrect. Therefore the odds on any positive result being valid are roughly even.

The frequency format could be easily digested by Hammerton's readers. You can "see" that the relative frequency is one out of two (i.e. 50%), and not 85%. Hammerton's subjects, however, were tested and failed on a single-event representation.

Here is a second example. In a fascinating article on mammography, Eddy (1982) reports that he asked 100 physicians questions of the following kind:

> The prevalence of breast cancer is 1% (in a specified population). The probability that a mammography is positive if a woman has breast cancer is 79%, and 9.6% if she does not. What is the probability that a woman who tests positive actually has breast cancer?____%

Eddy (1982) reports that 95 out of 100 physicians estimated the probability $p(cancer\,|\,positive)$ to be about 75%. However, if one applies Bayes' theorem to the information given, $p(cancer\,|\,positive)$ is only about 0.08 (or 8%). The judgment of these 95 physicians once more looks like an instance of the base-rate fallacy. College students, physicians, writers of medical textbooks (Eddy, 1982), and staff at the Harvard Medical School (Casscells, Schoenberger & Grayboys 1978) all seem to have equally great difficulties with problems of this kind. Reasoning about single-event probabilities (or percentages) does not seem to come naturally to them.

Let us now perform a thought-experiment with the mammography problem. Change the information representation in the mammography problem from single-event probabilities to frequencies:

> Imagine 100 people (think of a 10×10 grid). We expect that one woman has cancer and a positive mammography. Also, we expect that there are 10 more women with positive mammographies but no cancer. Thus we expect 11 people with positive mammographies. How many women with positive mammographies will actually have breast cancer?

With frequencies, you immediately "see" that only about *1 out of 11* women who test positive will have cancer. The base-rate fallacy disappears if the information is represented in frequencies. Note that by "frequencies", I mean

natural numbers. Let us now turn from the thought-experiment to real experiments.

Casscells et al. (1978) gave 60 staff and students of the Harvard Medical School the following problem, cast in *single-event probabilities* (except for the base rate):

> If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?

If one inserts these numbers into Bayes' theorem, the posterior probability that the person actually has the disease is 0.02 (assuming the test correctly diagnoses every person who has the disease—a piece of missing information).

Most of the staff and students at Harvard Medical School were hopelessly lost—almost half estimated this probability as 0.95 not 0.02. Only 11 participants answered 0.02. Note the amount of variability in the physicians' judgments about the probability of the disease! The modal answer of 0.95 was taken to be another instance of the *base-rate fallacy*, or *base-rate neglect*, as Tversky and Kahneman (1982) called it. The base rate of the disease (1/1000) is neglected, and judgment is based only (or mainly) on the characteristics of the test (here: the false positive rate). This seemed yet more proof of the stability of the base-rate fallacy.

But I will now apply to the Harvard Medical School problem the same frequency-representation procedure I applied to the preceding problems. If there is some kind of algorithm for statistical reasoning that works on frequency representations, changing the information representation in the Harvard Medical School problem from single-event probabilities and percentages to frequencies should make the base-rate fallacy disappear. Consequently, the large variability in judgments should also disappear.

Cosmides and Tooby (in press) have tested this prediction in a series of experiments with more than 400 Stanford undergraduates. They constructed a dozen or so variations of this medical problem, substituting step-by-step frequencies for single-event probabilities. In the original single-event version, the Stanford undergraduates gave almost the same low percentage of 0.02 answers as the staff and students at Harvard Medical School, 12% compared to 18% (Table 7.3). The original single-event version was somewhat ambiguous, because the true positive rate was not specified and Stanford undergraduates might not know what the term "false positive rate" means. Therefore, Cosmides and Tooby constructed a purified single-event version in which these ambiguities were eliminated:

> The prevalence of disease X is 1/1000. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the

> disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, 5% of all people who are perfectly healthy test positive for the disease.
>   What is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?____%

This made the percentage of 0.02 answers go up, but only to 36%. Rather dramatic effects were obtained, however, when the single-event format was changed into a frequency format. There were two major changes, the format of the *information* (first paragraph of the single-event version) and that of the task (second paragraph). To change the format of the information from single events to frequencies, (1) all probability information was expressed in frequencies such as "50 out of 1000" instead of 5%, and (2) a reference class ("Americans") was added on which these frequencies are defined. Nothing else was changed:

> One out of 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease.

To transform the *task* from estimating a single-event probability to estimating a frequency, a new sample of Americans was introduced, and the second paragraph of the single-event version was replaced by the following question:

> How many people who test positive for the disease will actually have the disease? ____out of____.

If our minds were not built to reason statistically, but only equipped with crude heuristics (consult Table 7.1), then the distinction between single events and frequencies should not matter. But it does. Table 7.3 shows how one can make almost everybody (or almost nobody, or any proportion inbetween) find the answer that corresponds exactly with the result of applying Bayes' theorem to the information given—that is, a probability of 0.02 or a frequency "1 out of 51," respectively. If both the information and the task were in terms of frequencies, this percentage was over seventy; if only one of the two was represented by frequencies, the percentage was inbetween the single-event and the frequency versions. If the frequency format was combined with asking the subjects to construct a pictorial frequency representation (i.e. to represent each person by a square, and mark those who do have the disease and those who test positive), then the percentage reached 92%.

Table 7.3 How to make the *base-rate fallacy* disappear: The Harvard Medical School problem (see Cosmides & Tooby, in press)

Representation of the problem	n	Answers in accordance with Bayes' rule (in %)
Original single-event format (Casscells et al., 1978)	60	18
Single-event format, replication	25	12
Information in frequency format, task in single-event format	25	56
Information in single-event format, task in frequency format	75	59
Information and task in frequency format	75	73
Information and task in frequency format, pictorial representation	25	92

Cosmides and Tooby's experimental variations, both in number and in detail, go beyond what I have described here. But my summary suffices to make the same point as with the conjunction fallacy. The conceptual distinction between single-event probabilities and frequencies seems to be as important for the untutored mind as it is for probability theory. It can make apparently stable cognitive illusions disappear.

These results have direct implications for teaching statistical reasoning.

7.2.4 Natural Sampling of Frequency Information

So far I have dealt with situations in which frequency information comes in one package, as in textbook problems or in newspapers. In many natural environments, and for animals or people in an illiterate world, however, frequencies must be *sequentially* learned through experience. How does an algorithm vary if we move from the standard single-event probability textbook problem to a corresponding ecological situation, in which the structure of the environment is sequentially learned through experience? Here is another thought experiment.

Let us transpose the above medical diagnosis problem to a non-literate society where physicians have to rely on their experience alone. Assume you are a physician. Your tribe has been afflicted for one year by a previously unknown and fatal disease. Everyone suspected of having the new disease is sent to you. You were lucky to discover one symptom that seems to signal the outbreak of the disease. What would it mean to be a Bayesian physician in this non-literate society?

You would encounter all information sequentially, as discrete cases that add up to frequencies. This information gathering is sometimes called *natural sampling* (Kleiter, 1993), a concept corresponding to Brunswik's (1955) *representative sampling*. So far you have seen 30 people suspected of having the disease. Ten of these turned out to have the disease, 20 did not. Of the 10 persons afflicted, 8 showed the symptom; of the 20 persons not afflicted, only 4 had the symptom. Now they bring in number 31. She has the symptom. What mental algorithm do you need in order to calculate the Bayesian posterior probability that she actually has the disease?

It turns out that in natural sampling this algorithm is quite simple—indeed, much simpler than required in those studies from which the base rate fallacy has been concluded. The algorithm needs only two absolute frequencies: the number a of people with symptom and disease, and the number b of people with symptom and no disease. These frequencies are $a = 8$ and $b = 4$, respectively. The algorithm to calculate the relative frequency $f(D | S)$ of people with disease D among those who have the symptom S is:

$$f(D \mid S) = \frac{a}{a+b} = \frac{8}{8+4}$$

If you are a Bayesian and want to calculate from the frequencies monitored so far the *posterior probability* $p(D | S)$ that patient number 31 has the disease, your mental algorithm is just as simple:

$$p(D \mid S) = \frac{8}{8+4}$$

Compare now this algorithm to that needed in the standard probability revision tasks of the heuristics and biases program. In the latter, the information is presented in terms of three single-event probabilities (forget for a moment the confusion between base rates and subjective priors): the prior probability $p(D)$, and the likelihoods $p(S|D)$ and $p(S| -D)$. For this representation of information, Bayes' theorem is:

$$p(D \mid S) = \frac{p(D)p(S \mid D)}{p(D)p(S \mid D) + p(-D)p(S \mid -D)}$$

The information (corresponding to the natural sampling condition) would be represented as $p(D) = 0.33$, $p(S | D) = 0.80$, and $p(S | -D) = 0.20$. Inserting these numbers into Bayes' theorem results in the following calculation:

$$p(D \mid S) = 0.33 \times 0.80/(0.33 \times 0.80 + 0.67 \times 0.20) = 0.67$$

The result is the same as in natural sampling, but the calculation is much more difficult.

The general point I want to make is that the way information is represented in an experiment, versus encountered in a natural environment, can require

reasoning algorithms of differing complexities. Even if these algorithms are mathematically equivalent, as they are in the thought experiment just presented, they can be computationally and psychologically different. Specifically, if information is encoded through natural sampling of frequencies—as opposed to laboratory studies which present three single-event probabilities—the following differences arise:

(1) In natural sampling, memory needs to monitor only *two* kinds of information, the frequencies *a* and *b*. *No* attention need be paid to the base rates themselves.
(2) In natural sampling, Bayes' rule reduces to a simple algorithm.
(3) Frequency information, naturally sampled, carries more information than single-event probabilities. Absolute frequencies contain information about the sample size (e.g. "3 out of 20", as opposed to $p = 0.15$), which allows for computing the precision (so-called second-order probabilities) of the information (Kleiter, 1993).

I know of very few studies that have used natural sampling instead of displaying three single-event probabilities. Christensen-Szalanski and Beach (1982) represented the information in a medical diagnosis problem (similar to those described earlier) both in the single-event probability format, as usual, and by natural sampling. In the single-event version the usual results were obtained, from which the base-rate fallacy has been concluded. In the natural sampling condition, subjects were shown 100 slides, one by one. Each slide contained information about one patient: whether or not the patient had pneumonia, and whether or not the test result was positive. As in the single-event version, the task was to estimate p(pneumonia | positive). The mean estimate in the natural sampling condition was 0.22, almost identical with the actual relative frequency f(pneumonia | positive) $= 6/(6 + 19) = 0.24$. (Although the means were very close, there was still considerable individual variability in estimates, perhaps due in part to individual differences in monitoring the actual frequencies.)

Here is a second study. One of the best publicized demonstrations of the base-rate fallacy outside of the realm of medical diagnosis problems is Tversky and Kahneman's (1982) *Cab problem*:

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

(i) 85% of the cabs in the city are Green and 15% are Blue.
(ii) a witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue rather than Green?

Tversky and Kahneman reported that the modal and median response of several hundred subjects was 0.80, whereas Bayes' theorem gives only 0.41. The median response is identical with the witness' hit rate—just as in some of the medical diagnosis problems—and this has been interpreted to mean that subjects neglect base rates. Bar-Hillel (1980; 1983) has tried many variations, such as presenting the base rates before or after the other information, and concluded that the base-rate fallacy was a robust phenomenon. Tversky and Kahneman (1980) suggested that the reason for this is that base rates tend not to be used unless they are seen as causal: "The proportions of Blue and Green cabs does not induce a differential propensity to be involved in accidents and this information is therefore neglected." (page 70).

Note that the problem, like those described earlier, is presented in terms of single-event probabilities and percentages. If base rates are neglected because they are not "causal", then the distinction between single-event probabilities and frequencies should not matter, since frequencies do not induce a "differential propensity to be involved in accidents", either.

Schlotterbek (1992) displayed the information in the Cab problem by means of natural sampling. In an analogy to the study by Christensen-Szalanski and Beach, 100 incidents of hit-and-run accidents were shown, one by one, using a computer display. In each case the subjects could see whether the cab was blue or green, and what the witness reported. After they had seen all 100 incidents, subjects estimated either the probability that a cab reported as "blue" is actually blue, in a new case, or the corresponding frequency f(blue | "blue"). Subjects were also asked (after they had seen all 100 cases) for their perceived four conjoint frequencies (blue cabs and report "blue", blue cabs and report "green", and so on). This allowed one to control for individual differences in perceived frequencies. Each subject's response to the frequency task was compared with the actual frequency, which was 12 out of 29, or 0.41, and with the corresponding individual frequency, calculated from the subject's reported conjoint frequencies.

Probability judgments corresponded well to the actual frequency (as in the Christensen-Szalanski & Beach study), and frequency judgments still better (median = .42; mean = .45). Half of the frequency judgments hit exactly either the actual frequency (12 out of 29) or the corresponding number calculated from subjects' perceived conjoint frequencies. Perceived conjoint frequencies were in very good correspondence with actual frequencies, with a slight overestimation of the smallest frequency (blue cabs and report "green") and underestimation for the largest (green cabs and report "green"). Neither of these two frequencies, however, is needed to solve the frequency task.

To summarize: the examples given, including sequential frequency processing, show that the distinction between single-event probabilities and frequencies is relevant to understanding how the mind reasons about a class of problems that are often termed *Bayesian probability revision* problems.

Thus far, we have seen how to make two cognitive illusions, the conjunction fallacy and the base-rate fallacy, largely disappear. I will now turn to a third prominent illusion.

7.2.5   How to Make Overconfidence Bias Disappear

Confidence in one's knowledge has been typically studied with questions of the following kind:

Which city has more inhabitants?

(a) Hyderabad
(b) Islamabad

How confident are you that your answer is correct?
       50%, 60%, 70%, 80%, 90%, 100%

Imagine you are an experimental subject. Your task is to choose one of these two alternatives. Suppose you choose Islamabad, as most subjects in previous studies did. Then you are asked to state your confidence, or subjective probability, that your answer "Islamabad" is correct. 50% confident means guessing; 100% confident means that you are absolutely sure that Islamabad is the larger city. From a large sample of questions, the experimenter counts how many answers in each of the confidence categories were actually correct.

The major finding of some two decades of research is the following (Lichtenstein, Fischhoff & Phillips, 1982): In all the cases where subjects said, "I am 100% confident that my answer is correct", the relative frequency of correct answers was only about 80%; in all the cases where subjects said, "I am 90% confident", the relative frequency of correct answers was only about 75%; when subjects said "I am 80% confident", the relative frequency of correct answers was only about 65%, and so on. Values for confidence were systematically *higher* than relative frequencies. This systematic discrepancy has been interpreted as an error in reasoning and has been named *overconfidence bias*. Quantitatively, overconfidence bias is defined as the difference between mean confidence and percentage correct.

Is overconfidence bias really a "bias" in the sense of a violation of probability theory? Let me rephrase the question: has probability theory been violated if *one's degree of belief (confidence) in a single event* (i.e. that a particular answer is correct) is different from the relative frequency of correct answers in the long run? From the point of view of the frequency interpretation, the answer is "no". In this view, probability theory is about frequencies; it does not apply to single-event judgments such as confidences. Therefore, no statement about confidences can violate the laws of probability. Even for Bayesians, however, the answer is not "yes". The issue here is not internal

consistency, but the relation between subjective probability and external (objective) frequencies, which is a more complicated issue and depends on conditions such as independence. In particular, if there is no feedback after each answer, as in this research, and if the true answers for a series of questions are dependent, one cannot expect that one's average degree of belief matches the relative frequency of correct answers. Consider, for instance, predictions of the following type: "Will there be snowfall on December 24, 1999, in downtown Boston? Yes/No." "Will there be snowfall on December 24, 1999, at Logan (Boston) airport? Yes/No." "Will there be snowfall on December 24, 1999, in Cambridge, Mass.? Yes/No." And so on. Assume, after careful consideration, that your probability that there will be snow is 0.7 in each case. Nevertheless, you cannot expect that your single-event confidences match the relative frequencies in the long run, because the outcomes are dependent. If it snows in downtown Boston, it will most likely snow in all places, and you appear to be underconfident; otherwise you will appear overconfident.

For these various reasons, a discrepancy between confidence in single events and relative frequencies in the long run should not be labeled simply an "error" in statistical and probabilistic reasoning, contrary to the claims in the heuristics-and-biases literature. It only looks that way from the perspective of a narrow interpretation of probability theory that blurs the fundamental distinction between single events and frequencies.

However, for the last two decades, many researchers have taken it for granted that any systematic difference between confidence and frequency is a reasoning error, a regrettable deviation from rationality. And they assumed that their task is to explain this discrepancy by some deficiency in our mental or motivational programming, such as a "confirmation bias" (Koriat, Lichtenstein & Fischhoff, 1980), "insensitivity to item difficulty" (von Winterfeldt & Edwards, 1986, page 128), and the tendency of humans in the Western world to overestimate their intellectual powers (Dawes, 1980). Similar to other "cognitive illusions", overconfidence bias has been suggested as an explanation for human disasters of many kinds, including deadly accidents in industry (Spettell & Liebert, 1986), errors in the legal process (Saks & Kidd, 1980) and systematic deviations from rationality in negotiation and management (Bazerman & Neale, 1986).

Many experiments have demonstrated the stability of the overconfidence phenomenon despite various "debiasing methods", such as warning subjects about overconfidence prior to the experiment or providing monetary incentives. We even used a bottle of French champagne as an incentive, but to no avail. Edwards and von Winterfeldt (1986, page 656) concluded in a tone of regret: "Can anything be done? Not much."

I will now apply the distinction between single-event probabilities and frequencies to the overconfidence bias. Take the same kind of general-knowledge

questions that have been used before to demonstrate the overconfidence bias. But now let our subjects make *frequency judgments*. After our subjects answered 50 general-knowledge questions of the Hyderabad–Islamabad type, in the usual format, they also had the opportunity to judge frequencies: "How many of these 50 questions do you think you answered correctly?"

If confidence in one's knowledge were truly biased due to confirmation bias, wishful thinking, or other deficits in cognition, motivation, or personality, then the difference between a single-event and a frequency representation should not matter. Overestimation should remain stable, as it does with warnings and bribes.

Table 7.4 shows the results of two experiments (Gigerenzer, Hoffrage & Kleinbölting, 1991). If one calculates, for each subject, the difference between mean confidence (averaged over 50 questions) and the relative frequency of correct answers, one finds as usual a stable positive difference that has been called the overconfidence bias: the value +13.8 is such a difference (multiplied by 100, and averaged across the 80 subjects in the first experiment). But the interesting issue is how the frequency estimates compare with the actual frequencies.

When we compared subjects' estimated frequencies with their true frequencies, there was *no* overestimation. Frequency judgments were quite accurate. In both experiments, the mean differences were even slightly negative, indicating a tendency towards underestimation. For instance, the figure −2.4 means that in the actual set of 50 questions, the estimated frequency of correct answers was, on the average, 1.2 lower than the true frequency. Subjects missed the true frequency by an average of only about 1 correct answer in a set of 50 questions.

Note that the very same subjects appear to be overestimating their knowledge, if one blurs the distinction between single-event probabilities and

Table 7.4 How to make the *overconfidence bias* disappear (see Gigerenzer, Hoffrage & Kleinbölting, 1991)

Difference between	Experiment 1 (n = 80)	Experiment 2 (n = 97)
Mean confidence and relative frequency of correct answers ("overconfidence bias")	+ 13.8	+ 15.4
Estimated frequency and frequency of correct answers	− 2.4	− 4.2

Note: To make values for frequency and confidence judgments comparable, all frequencies were transformed to relative frequencies. Values shown are differences multiplied by a factor of 100.

frequencies. You may think that this difference between single-event and frequency judgment is simply due to subjects having second thoughts about their performance at the end of the experiment. We have checked this. When the sequence "confidence judgments–frequency judgment" was repeated again and again (by presenting several sets of 50 questions in a sequence), subjects consistently gave different values for confidence and frequency.

This chapter is not the place to pursue the question of how to model these striking judgments. We have developed the theory of *probabilistic mental models* (Gigerenzer, Hoffrage & Kleinbölting, 1991), which explains this and related phenomena with an algorithm that infers both confidences and frequencies from frequency information—that is, from frequency information based on different reference classes.

To summarize: I have argued that the discrepancy between mean confidence and relative frequency of correct answers, known as "overconfidence bias", is not an error in probabilistic reasoning. It only seems that from a narrow normative perspective, in which the distinction between single-event probabilities and frequencies is blurred. If we ask our subjects about frequencies instead of single-event probabilities, we can make this stable phenomenon disappear. The conceptual distinction is much more effective with our subjects than money or French champagne.

The striking effect of frequency representations on apparent violations of probability theory, as reported in this chapter, seems to generalize to so-called violations of utility theory as well. For instance, Keren and Wagenaar (1987) showed that standard violations such as the "certainty effect" and the "possibility effect" (Kahneman & Tversky, 1979) largely disappear when a single gamble is changed into a repeated gamble (see also Keren, 1991; Lopes, 1981; Montgomery & Adelbratt, 1982). It also seems to generalize to a class of phenomena known as the "illusion of control" (Langer, 1975), which largely disappears if single-event estimates are replaced by judgments about a series of events (Koehler, Gibbs & Hogarth, in press).

7.3 CONCLUSIONS

Probability theory and psychology have historically been intertwined since the Enlightenment. The psychological theories of Locke, Hume, and Hartley provided the grounds for the classical interpretation of probability, in particular for the assumption that the mind unconsciously tallies frequencies and converts them into rational degrees of belief. This created the fiction of the reasonable man (*l'homme éclairé*) and the blurring of the distinction between objective frequencies and subjective probabilities. When, by the early nineteenth century, psychological theories had shifted to illusions, the reasonable man dissolved and the difference between frequencies and degrees of

belief became clear. The reasonable man gave way to the average man (*l'homme moyen*), and the frequency interpretation of probability emerged and became dominant. When the subjective interpretation—Bayesianism and subjective utility theory—regained influence in the second half of this century, these modern versions of Enlightenment rationality often did not distinguish between single-event probabilities and frequencies, nor between single and repeated gambles—just as classical probability theory had not. And many psychologists, following in these footsteps, also failed to make this distinction and found the human mind overflowing with cognitive illusions. Conflating single-event probabilities and frequencies now served the fiction of the irrational man.

Much of the current view is condensed in my economist colleague's dictum, "either reasoning is rational or it's psychological." Rationality is now defined in terms of formal algorithms or axioms, and psychology is called upon to explain the irrational. However, algorithms work on information, and information needs representation. To discuss rationality in terms of algorithms alone, good or bad ones, is incomplete if one does not pay attention to the kind of information representation that these algorithms were designed to work upon. Consequently, one cannot simply conclude from what looks like bad performance, or cognitive illusion, that there are poor algorithms. This *non sequitur* has been a basic flaw in the heuristics and biases program. When information is represented in terms of frequencies rather than single-event probabilities, apparently stable cognitive illusions tend to disappear.

ACKNOWLEDGEMENTS

REFERENCES

Bar-Hillel, M. (1980) The base rate fallacy in probability judgments. *Acta Psychologica*, 44, 211–33.

Bar-Hillel, M. (1983) The base-rate fallacy controversy. In R.W. Scholz (ed.), *Decision making under uncertainty*. North-Holland, Amsterdam.

Bazerman, M.H. & Neale, M.A. (1986) Heuristics in negotiation. In H.R. Arkes & K.R. Hammond (eds.), *Judgment and decision making: an interdisciplinary reader*. Cambridge University Press, Cambridge, pages 311–21.

Bernoulli, J. (1713) *Ars conjectandi*. Basel.

Birnbaum, M.H. (1983) Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, 96, 85–94.

Brehmer, B. & Joyce, C.R.B. (eds.) (1988) *Human judgment: The SJT view.* North-Holland, Amsterdam.

Brunswik, E. (1955) Representative design and probabilistic theory in a functional psychology. *Psychological Review,* 62, 193–217.

Casscells, W., Schoenberger, A. & Grayboys, T. (1978) Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine,* 299, 999–1000.

Christensen-Szalanski, J.J.J. & Beach, L.R. (1982) Experience and the base-rate fallacy. *Organizational Behavior and Human Performance,* 29, 270–8.

Condillac, E. (1754/1793) *Traité des sensations.* In *Oeuvres,* Vol. 3, Paris.

Cosmides, L. & Tooby, J. (1992) Cognitive adaptations for social exchange. In J.H. Barkow, L. Cosmides & J. Tooby (eds), *The Adapted Mind.* Oxford, Oxford University Press, pp. 163–228.

Cosmides, L. & Tooby, J. (in press) Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition.*

Cournot, A.A. (1843) *Exposition de la théorie des chances et des probabilités.* Paris.

Danziger, K. (1990) *Constructing the subject: Historical origins of psychological research.* Cambridge University Press, Cambridge.

Daston, L.J. (1988) *Classical probability in the Enlightenment.* Princeton University Press, Princeton, NJ.

Daston, L.J. (1992) The doctrine of chances without chance: Determinism, mathematical probability, and quantification in the seventeenth century. In M.J. Nye et al. (eds.), *The Invention of Physical Science.* Kluwer Academic Publishers, Netherlands.

Dawes, R.M. (1980) Confidence in intellectual judgments vs. confidence in perceptual judgments. In E.D. Lantermann & H. Feger (Eds.), *Similarity and choice: Papers in honor of Clyde Coombs.* Huber, Berne, pages 327–45.

Eddy, D.M. (1982) Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic & A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge, pages 249–67.

Edwards, W. (1968) Conservatism in human information processing. In B. Kleinmuntz (ed.), *Formal Representation of Human Judgment.* Wiley, New York.

Edwards, W., Lindman, H. & Savage, L.J. (1963) Bayesian statistical inference for psychological research. *Psychological Review,* 70, 193–242.

Edwards, W. & von Winterfeldt, D. (1986) On cognitive illusions and their implications. In H.R. Arkes & R.R. Hammond (eds.), *Judgment and decision making: An interdisciplinary reader.* Cambridge University Press, Cambridge, pages 642–79.

Fiedler, K. (1988) The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research,* 50, 123–9.

Fischhoff, B. (1988) Judgment and decision making. In R.J. Sternberg & E.E. Smith (eds.), *The Psychology of Human Thought.* Cambridge University Press, Cambridge, pages 153–87.

Gallistel, C.R. (1990) *The Organization of Learning.* MIT Press, Cambridge, MA.

Gigerenzer, G. (1991) How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology,* 2, 83–115.

Gigerenzer, G. (1993) The Superego, the Ego, and the Id in statistical reasoning. In G. Keren & C. Lewis (eds.), *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues,* Erlbaum, Hillsdale, NJ, pages 313–39.

Gigerenzer, G., Hell, W. & Blank, H. (1988) Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance,* 14, 513–25.

Gigerenzer, G. Hoffrage, U. & Kleinbölting, H. (1991) Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–28.

Gigerenzer, G. & Hug, R. (1992) Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127–71.

Gigerenzer, G. & Murray, D.J. (1987) *Cognition as Intuitive Statistics.* Erlbaum, Hillsdale, NJ.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. & Krüger, L. (1989) *The Empire of Chance: How Probability Changed Science and Everyday Life.* Cambridge University Press, Cambridge.

Gould, S.J. (1992) *Bully for brontosaurus. Further Reflections in Natural History,* Penguin Books, London.

Hacking, I. (1975) *The Emergence of Probability.* Cambridge University Press, Cambridge.

Hacking, I. (1990) *The Taming of Chance.* Cambridge University Press, Cambridge.

Hammerton, M. (1973) A case of radical probability estimation. *Journal of Experimental Psychology,* **101,** 252–4.

Hartley, D. (1749) *Observations on Man, His Frame, His Duty, and His Expectations.* London.

Hasher, L., Goldstein, D. & Toppino, T. (1977) Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior,* **16,** 107–112.

Hasher, L. & Zacks, R.T. (1979) Automatic and effortful processes in memory. *Journal of Experimental Psychology: General,* **108,** 356–88.

Hertwig, R. & Gigerenzer, G. (1994) The chain of reasoning in the conjunction task. Unpublished manuscript.

Hume, D. (1739/1975) *A Treatise of Human Nature.* Clarendon Press, Oxford.

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

Kahneman, D., & Tversky, A. (1973) On the psychology of prediction. *Psychological Review,* **80,** 237–51.

Kahneman, D., & Tversky, A. (1979) Prospect theory: An analysis of decision under risk. *Econometrica,* **47,** 263–91.

Keren, G. (1991) Additional tests of utility theory under unique and repeated conditions. *Journal of Behavioral Decision Making,* **4,** 297–304.

Keren, G. & Wagenaar, W. A. (1987) Violation of utility theory in unique and repeated gambles. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **13,** 387–91.

Kleiter, G.D. (1993) Natural sampling: Rationality without base rates. In G.H. Fischer & D. Lanning (eds.), *Contributions to Mathematical Psychology, Psychometrics, and Methodology.* Springer, New York, pages 369–82.

Koehler, J.J., Gibbs, B.J. & Hogarth, R.M. (in press) Shattering the illusion of control: Multi-shot versus single-shot gambles. *Journal of Behavioural Decision Making.*

Kolmogoroff, A.N. (1933) Grundbegriffe der Wahrscheinlichkeitsrechnung. *Ergebnisse der Mathematik,* 2(3), 196–262.

Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980) Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory,* **6,** 107–18.

Krüger, L., Daston, L. & Heidelberger, M. (eds.) (1987) *The Probabilistic Revolution. Vol. I: Ideas in History.* MIT Press, Cambridge, MA.

Krüger, L., Gigerenzer, G. & Morgan, M.S. (eds.) (1987) *The Probabilistic Revolution. Vol. II: Ideas in the Sciences,* MIT Press, Cambridge, MA.

Langer, E.J. (1975) The illusion of control. *Journal of Personality and Social Psychology,* **32,** 311–28.

Lichtenstein, S., Fischhoff, B. & Phillips, L.D. (1982) Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (eds.), *Judgment under uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, pages 306–34.

Lopes, L.L. (1981) Decision making in the short run. *Journal of Experimental Psychology: Human Learning and Memory*. 7, 377–85.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, San Francisco.

Mises, R.v. (1957) *Probability, Statistics, and Truth*, Allen and Unwin, London.

Montgomery, H. & Adelbratt, T. (1982) Gambling decisions and information about expected value. *Organizational Behaviour and Human Performance*, 29, 39–57.

Neyman, J. (1977) Frequentist probability and frequentist statistics. *Synthese*, 36, 97–131.

Nisbett, R.E. & Ross, L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgment*. Prentice-Hall, Englewood Cliffs, NJ.

Piaget, J. & Inhelder, B. (1975) *The Origin of the Idea of Chance in Children*, Norton, New York, (original work published in 1951).

Piattelli-Palmarini, M. (1991) Probability blindness. Neither rational nor capricious. *Bostonia*, March/April, pages 28–35.

Poisson, S-D. (1837) *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris.

Porter, T.M. (1986) *The rise of statistical thinking, 1820–1900*. Princeton University Press, Princeton.

Quetelet, L.A.J. (1835) *Sur l'homme et le développement de ses facultés, ou essai de physique sociale*. Bachelier, Paris.

Real, L.A. & Caraco, T. (1986) Risk and foraging in stochastic environments: Theory and evidence. *Annual Review of Ecological Systems*, 17, 371–90.

Saks, M., & Kidd, R.F. (1980) Human information processing and adjudication: Trial by heuristics. *Law and Society Review*, 15, 123–60.

Savage, L.J. (1954) *The Foundations of Statistics*. Wiley, New York.

Schlotterbek, M. (1992) *Bayesian Inference Without Base Rates*. Diploma thesis, University of Tübingen, Germany.

Schum, D. (1990) Discussion. In R.M. Hogarth (ed.), *Insights in Decision Making* University of Chicago Press, Chicago, pages 217–33.

Shafer, G. (1989) *The Unity and Diversity of Probability*, Inaugural lecture, 20 November 1989, University of Kansas.

Shapiro, B.J. (1983) *Probability and Certainty in Seventeenth-century England*. Princeton University Press, Princeton.

Slovic, P., Fischhoff, B. & Lichtenstein, S. (1976) Cognitive processes and societal risk taking. In J.S. Carrol & J.W. Payne (eds.), *Cognition and Social Behavior*. Erlbaum, Hillsdale, NJ.

Spettell, C.M. & Liebert, R.M. (1986) Training for safety in automated person-machine systems. *American Psychologist*, 41, 545–550.

Stigler, S.M. (1986) *The History of Statistics: The Measurement of Uncertainty before 1900*. Belknap Press of Harvard University Press, Cambridge.

Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–31.

Tversky, A. & Kahneman, D. (1980) Causal schemas in judgments under uncertainty. In M. Fishbein (eds.), *Progress in Social Psychology*. Erlbaum, Hillsdale, NJ, pages 49–72.

Tversky, A. & Kahneman, D. (1982) Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, pages 153–60.

Tversky, A. & Kahneman, D. (1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293–315.

Winterfeldt, D. von, & Edwards, W. (1986) *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge.