In K. Smith, J. Shanteau & P. Johnson (Eds.), *Psychological investigations of competence in decision making* (pp. 188–225). Cambridge, UK: Cambridge University Press. (2004)

7 Representations of Uncertainty and Change: Three Case Studies with Experts

Elke M. Kurz-Milcke, Gerd Gigerenzer, and Ulrich Hoffrage

Otto Neurath, a driving force behind the Vienna Circle's scientific worldview, once remarked that "it is generally not a good sign when scholars concern themselves too much with the foundations and the history of their discipline, instead of working to find new and exact statements about the topic they are investigating" (1930/1931, p. 107, our translation). Consequently, Neurath thought it advisable to limit such concerns to an occasional "Sunday."

We have not followed Neurath's advice. The three case studies we are presenting have not only benefited from historical research (not necessarily on Sundays), but to some extent owe their existence to it. In our experience, certain aspects of current practice become apparent only in historical perspective; we may study past practices in order to come to an improved understanding of current practices. Hence, in our view, an historical and a scientific perspective need not be rival siblings competing for privileged attention, as they seem to be in Neurath's worldview.

Representational Practices

In this chapter we focus on one particular aspect of practice, namely, experts' representational practices. Experts and, a fortiori, laypeople do

Case Studies I and II were supported by the Deutsche Forschungsgemeinschaft (Ho 1847/1-1). Case Study III was supported in part by a scholarship of the American Psychological Association to the first author and was part of her dissertation research at Bowling Green State University, Ohio. The first author is grateful to Ryan D. Tweney for his advice and support of the research presented in Case Study III. We thank Krishna Bharathi, Seth Bullock, Adam Goodie, Martin Lages, and Laura Martignon for helpful comments and Jill Vyse for her careful editing of the manuscript.

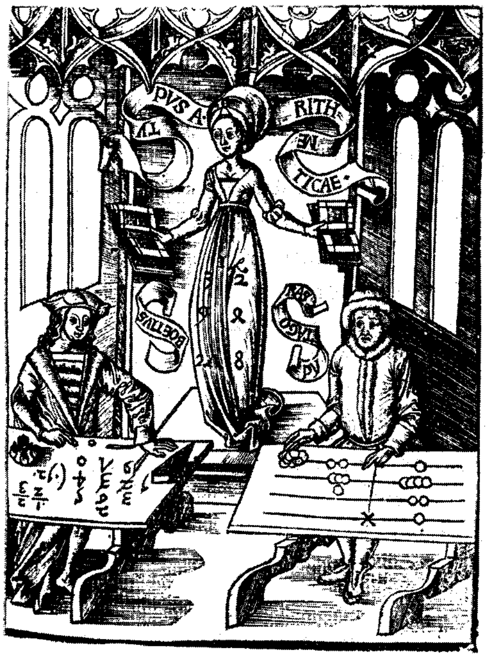

Figure 7.1 Two ways of doing arithmetic. "Dame Arithmetic" from Gregor Reisch, *Margarita Philosophica*, Strassbourg, 1504. (From *The Abacus*, p. 25, by Parry Moon, 1971, New York: Gordon and Breach. Reprinted with permission.)

not always consciously choose a representation, but rely on one that has been established by previous practice and convention. We have chosen a print from the early 16th century to illustrate what we mean by representational practices (Figure 7.1). On the right side of this print, Pythagoras is depicted operating on a calculating board. The left side shows Boethius, one of the earliest and most widely read scholastics, calculating with the symbols of the "new arithmetic." Note their facial expressions.

Dame Arithmetic, in the background, has obviously chosen her favorite. Anyone who has tried division on an abacus, which is but a slight modification of a calculating board, will probably agree with Dame Arithmetic's preference for Arabic numerals and computational symbols that can be represented in writing. However, the new arithmetic does not provide, for all purposes and in all contexts, *the* best representation. Even in our century, abacists have consistently outperformed contestants using calculating machines, at least in terms of speed (Dilson, 1968). On the other hand, the mathematical developments that we refer to later, probability calculus and differential calculus, would not have been possible, as we know them, without written numerical and operational symbols.

One of the criteria by which the usefulness of a particular representation can be judged is computational ease. Consider, for example, a comparison of the Arabic and Roman number systems. Addition and subtraction are easier with Roman numerals (for computation, IIII should be used rather than IV), whereas multiplication and division are much easier with Arabic numerals (see Norman, 1993; Zhang & Norman, 1995, for a detailed comparison of various numeration systems). Furthermore, a representation can help or hinder insight, or highlight certain aspects and make others less easy to see. It is therefore not surprising that at various times mathematical representations have been vehicles of political change. One such occasion was the legal adoption of the metric system in France in 1795 under the motto "for all the people, for all times"; in spite of the motto, "the people" on the streets and in the market squares kept up their old, tried and tested representational practices for many years afterward.

In our case studies, we were interested in the ways in which experts – acquired immune deficiency syndrome (AIDS) counselors, physicians, natural scientists, and mathematicians – represent information in order to draw quantitative inferences. In Case Studies I and II, experts associated with the medical field, AIDS counselors and physicians, had to infer the probability of a disease when a positive test was obtained. The representations employed by the counselors in Case Study I were strikingly uniform and remarkably inefficient, for the experts as well as for their clients (Gigerenzer, Hoffrage, & Ebert, 1998). Case Study II demonstrates an effective way of improving diagnostic reasoning, by using a different representational format (Hoffrage & Gigerenzer, 1998). In Case Study III, experts associated with different academic disciplines (chemistry, physics, mathematics) were asked to solve a problem requiring a differential equation for its exact solution. These scientists relied on very different representations to solve a mixture problem (Kurz, 1997).

Expertise is often discussed in terms of how much information an expert has, and in terms of his or her competence in selecting and processing the relevant information. In this chapter, we argue that experts' reasoning does not simply occur inside the experts' heads, but is performed to a substantial degree by the external representation of information the expert chooses or relies on. Shanteau (1992) has argued that experts' performance cannot be described generically, that is, without taking task characteristics into account. Our argument is in line with his. Representation is a task characteristic. Our case studies involve experts' use of a calculus, of the probability calculus in Case Studies I and II, and of the differential calculus in Case Study III. Finding adequate representations, and also notations was, in fact, central to the inception and the historical development of these calculi.

Calculi

The Latin term *calculus* means "little stone" and has present-day derivatives in the English words *calculate* and *calculator*. Calculi, or pebbles, were among the earliest calculation and bookkeeping aids (Damerow, 1995), leading to devices like the counting board, which was used throughout the Middle Ages, and the abacus (or soroban), which is still used today, especially in Asia (Moon, 1971). Nowadays we do not think of a calculus as a material object (except, perhaps, when the term is used in the medical context), but rather as a set of formalisms to solve problems concerning, for instance, uncertain events and changing phenomena. This modern sense of the term stems from the 17th century, and in this sense it relates to the inception of differential and integral calculus (Bos, 1993; Grattan-Guinness, 1980) and of the calculus of probability (Daston, 1988; Gigerenzer et al., 1989).

In the 17th and 18th centuries, the newly developed calculi of change and uncertainty had no existence apart from their subject matters. The same people who worked on problems that we today would consider to be applied problems also concerned themselves with the analytic aspects of the mathematics (Bos, 1993, p. 118). Moreover, 18th-century mathematics was dominated by *mixed mathematics*, a category that as such has ceased to be familiar to us. Mixed mathematics subsumed the study of topics that we today would consider to be separate fields of study, if not disciplines, for example, navigation, architecture, or geography. But no matter which kind of 18th-century mathematics is concerned, "all of mathematics, including pure mathematics, studied some*thing*" (Daston, 1988, p. 54). Even today, insofar as mathematical formalisms are tied to specific representations, be they graphic, verbal, or mental, mathematics retains a material aspect.

A calculus consists of a fund of basic rules as well as basic concepts. Thus, a calculus incorporates semantics. G. W. Leibniz's program of a universal calculus, formulated at the dawn of the Enlightenment, illustrates this point very nicely. Leibniz envisioned an encompassing system of signs – one for every basic concept – and rules to combine them and its use to record all human knowledge. Then, according to Leibniz's program, it would be possible to settle disputes by computation and, similarly, to acquire new insights based on the application of this calculus. What a daring vision! As Leibniz envisioned it, this formal system was not an end in itself but was meant to serve scientific progress. The calculus's final success, of course, hinged on an important precondition: universal agreement as to its appropriateness. Unsurprisingly, this universal calculus did not materialize on the envisioned scale (which does not lessen Leibniz's computational achievements, which were extraordinary, including, among others, the inception of the differential and integral calculus, contributions to the probability calculus and to logic, as well as the construction of the first calculating machines capable of performing the four basic arithmetic operations).

Leibniz's universal calculus has remained a dream, and the term calculus has lost much of its Leibnizian grandeur. Today, the term is mostly used to refer to particular formal systems, as, for example, the differential calculus or the probability calculus. But even these calculi are not what Leibniz had envisioned, that is, unitary formal systems. Historical scholarship has taught us that these calculi were shaped by alternative, at times competing, proposals. We argue that there are always multiple alternative representations and interpretations in a calculus, and that this variability in representational practice can be functional, allowing a calculus to be relevant for various tasks and in various domains. Our case studies of present-day experts and their representational practices show how this plurality of representation can be a resource. Our case studies are preceded by historical "prisms" that intend to make the spectra of interpretations and representations of the calculi visible.

Representing Uncertainty

Observing Historical Spectra: The Calculus of Uncertainty

According to legend, the calculus of uncertainty is one of the few seminal ideas that has an exact birthday. In 1654, the now famous correspondence between Blaise Pascal and Pierre Fermat first cast the calculus of probability in mathematical form. Ian Hacking (1975) argued that this probability, which 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 the death of kings); the other face was epistemic, concerned with degrees of belief or opinion warranted by authority. In his view, the "20th-century" duality between objective frequencies and subjective probabilities existed then as now. Barbara Shapiro (1983) and Lorraine Daston (1988), however, have argued that probability in the 17th and 18th 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 die 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 of the guilt of the accused); verisimilitude and epistemological modesty, among others. Over the centuries, probability also conquered new territories and created further meanings, such as in quantum physics, and lost old territory, such as the probability of causes (Daston, 1988).

The important point is that the calculus of probability began with several interpretations, and this plurality is still with us. This does not mean that the relationship between these interpretations has remained stable – on the contrary. For instance, the two major faces of probability, subjective belief and objective frequencies, began as equivalents and ended up as diametric opposites. For Jakob Bernoulli and the other Enlightenment mathematicians, belief and frequencies were just two sides of the same coin, and the ease with which the Enlightenment probabilists slid from one interpretation to the other is breathtaking – from today's point of view. Poisson eventually distinguished subjective belief and objective frequencies, and the political economist and philosopher Antoine Cournot (1843/1975) seems to have been the first to go one step further and eliminate subjective belief from the subject matter of mathematical probability: Mathematical probability was not a measure of belief. There is a broader intellectual and social context in which the demise of subjective belief as the subject matter of probability is embedded. The French Revolution and its aftermath shook the confidence of the mathematicians in the existence of a single shared standard of reasonableness. The consensus and the values of the intellectual and political elites fragmented, and degrees of belief became associated with wishful thinking and irrationality. By that time, the calculus of probability had lost its subject matter, the judgment and decision making of reasonable people (Gigerenzer et al., 1989, ch. 1).

By 1840, the calculus of uncertainty was no longer about mechanical rules of rational belief embodied in an elite of reasonable men, but about the observable properties of the *average man* (*l'homme moyen*), the embodiment of mass society, if not of mediocrity. Adolphe 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 them 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 or crimes in Paris or in London, proved to be strikingly stable over the years, and this was cited as evidence that moral phenomena are governed by the laws of a society rather than by the free decisions of its members. In 19th-century France, statistics became known as *moral science*. Quetelet offered a model of human behavior as erratic and unpredictable at the individual level, but governed by statistical laws and predictable at the level of society. This model 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, through analogy with the statistical laws of society, physics was revolutionized.

Throughout most of the 19th and 20th centuries, the *probabilistic revolution* (Krüger, Daston, & Heidelberger, 1987; Krüger, Gigerenzer, & Morgan, 1987) was about frequencies, not about degrees of belief: from the kinetic theory of gas to quantum statistics, and from population genetics to the Neyman–Pearson theory of hypothesis testing. As is well known, subjective probability regained acceptance in the second half of the 20th century with the pioneering work of Bruno de Finetti and Frank Ramsay in the 1920s and 1930s and of Leonard Savage in the 1950s. The reasonable man, once exiled from probability theory, made 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 (whose most prominent species are 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 dominate theoretical economics and artificial intelligence. The territory has been divided up. As Glenn Shafer (1989) complained, "conceptually and institutionally, probability has been balkanized" (p. 15).

To summarize: Since its inception, the calculus of uncertainty has had not one subject matter, but a multitude, that is, there has always been more than one interpretation of this calculus. The two most prominent ones are objective frequencies and subjective degrees of belief. The important point for this chapter is that each of these two interpretations is linked with a specific class of representations. Observed frequencies, for instance, can be represented by discrete elements that are the final tally of a counting process, and which are different from degrees of belief and single-event probabilities. The two representations we will focus on are *natural frequencies* and *probabilities*. The former are pure observed frequencies; the latter are the typical representations of degrees of belief.

Experts' Representations of Uncertainty. In this section, we summarize two case studies that demonstrate how the representation of statistical information – single-event probabilities and natural frequencies – affects human reasoning in Bayesian inference tasks. We first give an example to show how the Bayesian solution can be derived from either of the two representations. Then we report on how AIDS counselors reason in such a task and how they represent the relevant information spontaneously (Case Study I). Finally, we show how performance can be considerably improved by altering the representation of information (Case Study II).

Task Analysis. Consider the situation of a young heterosexual man who has undergone a human immunodeficiency virus (HIV) test. He does not engage in activities considered risky, such as intravenous (IV) drug use or homosexual practices. Yet, the result – after repeatedly applying the enzyme-linked immunosorbent assay (ELISA) and the Western blot test – comes back positive. What is the probability that he actually has HIV? If the test is positive, the probability of being infected – also known as the *positive predictive value* (PPV) – can be computed by Bayes's rule:

$$\text{PPV} = \frac{p(\text{HIV})\,p(\text{pos}|\text{HIV})}{p(\text{HIV})\,p(\text{pos}|\text{HIV}) + p(\text{no HIV})\,p(\text{pos}|\text{no HIV})}, \quad (1)$$

where p(HIV) denotes the prevalence of HIV in the respective population, p(no HIV) equals $1 - p$(HIV), p(pos|HIV) denotes the sensitivity of the test, and p(pos|no HIV) denotes the false positive rate of the test. Often the specificity rather than the false positive rate of a test is reported; the specificity of a test equals $1 - p$(pos|no HIV). To compute the PPV, we consulted the literature for estimates of these probabilities. The prevalence of HIV in heterosexual men with no known risk factors is estimated to be 0.01%. The best estimates for the sensitivity and specificity of the respective testing procedure (repeated ELISA and Western blot testing) is 99.8% for the sensitivity of the test and 99.99%

for its specificity (see Gigerenzer et al., 1998). Inserting these values into Bayes's rule results in a PPV of .50, or 50%, which is consistent with reports in the literature (Deutscher Bundestag, 1990, p. 121; Stine, 1996, p. 338).

Do people understand Bayes's rule and can they infer the PPV from a given prevalence, sensitivity, and false positive rate? There is considerable empirical evidence that suggests the answer is "no" (e.g., Casscells, Schoenberger, & Grayboys, 1978; Eddy, 1982; Gigerenzer & Hoffrage, 1995). The good news is that this negative answer need not be the cause of utter pessimism for the following reason. Note that the statistical information in our HIV example was represented in terms of probabilities. This probability-based format is the information format generally used in medical textbooks and curricula, as well as in the experiments that have demonstrated people's (including physicians') poor performance in Bayesian inference tasks. However, as we saw in the preceding section, probabilities constitute only one way of representing the relevant information; it can also be represented in terms of *natural frequencies*, that is, the absolute frequencies that result from observing cases that have been representatively sampled from a population.

Unlike probabilities, natural frequencies are not normalized with respect to the base rates of disease or no disease. Using natural frequencies, computation of the PPV can be communicated as follows: "Imagine that 10,000 heterosexual men are tested. One has the virus, and he will with practical certainty test positive (sensitivity = 99.8%). Of the remaining uninfected men, one will also test positive (false positive rate = 0.01%). This means that we expect that two men will test positive, and only one of them has HIV. Thus, the chance of having the virus given a positive test is one out of two, or 50%." In general, the PPV is the number of true positives (TP) divided by the number of true positives and false positives (FP):

$$\text{PPP} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{2}$$

Figure 7.2 illustrates the fact that Bayesian computations are simpler with natural frequencies than with probabilities or percentages, where the relevant statistical information is inserted in Equation 1 (left-hand side) and Equation 2 (right-hand side).

To compute the formula on the right-hand side, fewer cognitive operations need to be performed. Comparing the two equations yields

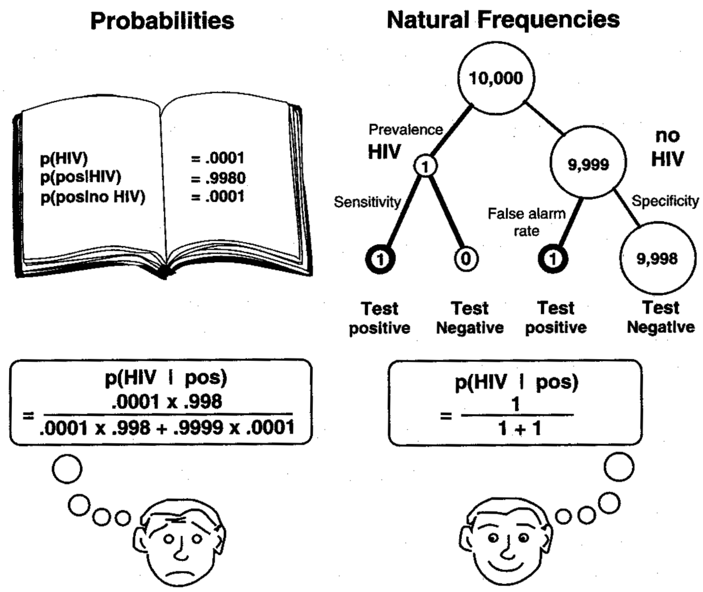

Figure 7.2 Two ways of inferring the predictive value of a positive HIV test. Relevant information is represented either as probabilities or as natural frequencies.

two further (related) results. First, with frequency representation only two pieces of information, the symptom and disease frequencies and the symptom and no disease frequencies (the two bold circles in Figure 7.2), need to be used. Second, and as a consequence of this, the base rate frequency (1 out of 10,000) can be ignored. On balance, in this kind of Bayesian inference task we obtain the same result with less information and fewer cognitive operations. We have chosen our favorite!

As a final example, let us compute the risk of a young homosexual man having HIV, given a positive test result. Assuming that the prevalence for this group is 1.5%, try to compute the PPV by using the probability representation first. (To make a fair comparison, you should think of how you would have dealt with this task before reading the previous pages of this chapter.) Now consider the frequency representation: Out of 10,000 homosexual men, about 150 have the virus and they will probably all test positive. Of the remaining uninfected men, one will also test positive. Thus, we expect that 151 men will test positive and that 150 of these men have HIV. Given a positive test result, the chance of having the virus is therefore 150 out of 151, or 99.3%.

In this way, representing information in terms of natural frequencies can help us "see" the correct Bayesian answer. The numbers can be adjusted; the point is that these numbers have to be represented somehow and that the representation chosen makes the computation more or less easy.

AIDS Counseling for Low-Risk Clients (Case Study I)

How do AIDS counselors communicate the meaning of a positive test result in actual counseling sessions? Do they communicate the risk in probabilities or natural frequencies? The study by Gigerenzer et al. (1998) on AIDS counseling in German public health centers seems to be the only study to have investigated what AIDS counselors tell a low-risk client about the meaning of a positive test. One of the authors of the study visited 20 counseling locations as a client wishing to take an HIV test. He asked the counselor the following questions in the order indicated (unless the counselor provided the information unprompted):

1. Sensitivity of the HIV test	If one is infected with HIV, is it possible to have a negative test result? How reliably does the test identify a virus if the virus is present?
2. Specificity of the HIV test	If one is not infected with HIV, is it possible to have a positive test result? How reliable is the test with respect to a false positive result?
3. Prevalence of HIV in heterosexual men	How frequent is the virus in my risk group, that is, heterosexual men, 20 to 30 years old, with no known risk – such as IV drug use?
4. Predictive value of a positive test	What is the probability that men in my risk group actually do have HIV after a positive test?
5. Window period	How much time has to pass between infection and test, so that antibodies can be detected? (For the present purpose, we will omit the results concerning this question.)

The interview included the following scheme for clarifying questions. If the counselor's answer was a quantitative estimate (a figure or a range), or if the counselor said that he or she could not (or did not want to) give a more precise answer, then the client went on to the next question. If the answer was qualitative (e.g., "fairly certain"), or if the counselor misunderstood or avoided answering the question, the client asked for further clarification and, if necessary, repeated the request for clarification once. If, after the third attempt, there was still no success, the client did not push further and went on to the next question. When the client needed to ask for clarification concerning the prevalence of HIV (Question 3), he always repeated his specific risk group; when asking for clarification concerning the PPV (Question 4), he always referred to the specific prevalence in his risk group, namely, "heterosexual men, 20 to 30 years old, with no known risk-related activity such as IV drug use." If the counselor asked for more information, which happened in only 11 of the 20 sessions, the client explained that he was 27 years old and monogamous, and that neither his current nor his (few) previous sexual partners used IV drugs or engaged in other risky behavior. In 2 of these 11 cases, the client was given a detailed questionnaire to determine his risk; in 1 of these 2 cases, the counselor did not look at the questionnaire and the client still had it in his hands when he left the location. Before turning to the general results of this study, we present excerpts from four typical counseling sessions. The client's questions are abbreviated (e.g., Sensitivity?) and followed by the information provided by the counselors. The counselors' answers to the client's clarifying questions are preceded by a hyphen.

Session 1: A Social Worker

Sensitivity? False negatives really never occur. Although if I think about the literature, there were reports about such cases. – I don't know exactly how many. – It happened only once or twice. *False Positives?* No, because the test is repeated; it is absolutely sure. – If there are antibodies, the test identifies them unambiguously and with absolute certainty. – No, it is absolutely impossible that there are false positives; by repeating the test it is absolutely certain. *Prevalence?* I can't tell you this exactly. – Between about 1 of 500 and 1 of 1,000. *Positive predictive value?* As I have now told you repeatedly, the test is absolutely certain.

Session 2: A Physician

Sensitivity? When there are enough antibodies, then the test identifies them in every case. Two tests are performed; the first test is in its fourth generation and is tuned to be very specific and sensitive. Nevertheless it is tuned in a way that it is more likely to identify positives than negatives. – 99.8% sensitivity and specificity. But we repeat the test, and if it comes out positive, then the result is as solid as cast iron. *False Positives*? With certainty, they don't occur; if there are false results, then only false negatives, occurring when the antibodies have not formed. – If you take the test here, including a confirmatory test, it is extremely certain. In any case the specificity is 99.7%. This is as solid as cast iron. We exclude confusions by using two tests. *Prevalence*? The classification of individuals into risk groups is by now outdated, therefore one cannot look at this that way. – I don't remember this. There is a trend for the virus to spread in the general public. Statistics are of no use for the individual case! *Positive predictive value*? As I already have said: extremely certain, 99.8%.

Session 3: A Physician

Sensitivity? The test is very, very reliable, that is, about 99.98%. *False Positives*? The test will be repeated. After the first test, we do not speak of positive, but only of reactive. When all tests are performed, then the result is sure. – It is hard to say how many false positives occur. – How many precisely? I would have to look up the literature to see if I could find this information there. *Prevalence*? That depends on the region. – Of the approximately 67,000 infected people [in Germany], 9% are heterosexual. – In Munich we have 10,000 infected people, that is, 1% of the population. But these are only numbers, which tell you nothing about whether you have the virus or not. *Positive predictive value*? As I have already have mentioned, the result is 99.98% sure. If you get a positive result, you can trust it.

Session 4: A Social Worker

Sensitivity? Very, very reliable. – No, not absolutely sure, such a thing doesn't exist in medicine, because it may be possible that the virus cannot be identified. – Close to 100%; I don't know exactly. *False Positives*? They exist, but are extremely rare. – On the order of one tenth of 1%. Probably less. However, in your risk group, compared with high-risk groups, false positives are proportionally more frequent. – I don't know the exact value. *Prevalence*? With the contacts you have had, an infection is unlikely. – Generally one can't say. In our own institution, among some 10,000 tests in the last 7 years, there were only three or four heterosexuals, nondrug addicts, or similar non-risk-group persons who tested positive. *Positive predictive value?* As mentioned, the test is not 100% sure. If the test confuses the [HIV] antibodies with others, then other methods such as repeated tests do not help. And if someone like you does not have a real risk, then I could imagine that even 5% to 10% of those who get a positive result will have gotten a false positive result.

How Did the Counselors Represent Statistical Information? The client was provided with information concerning sensitivity by 19 out of 20 counselors. (One of them refused to give any information concerning sensitivity, specificity, or predictive value before the test result was obtained. When the client collected the test result, he didn't receive any information either.) Most counselors gave the client realistic information concerning sensitivity (Table 7.1). However, 5 out of the 19 counselors incorrectly informed the client that even after the window period, it would be impossible to get a false negative result.

The client was informed incorrectly that false positives do not occur by 13 out of 19 counselors (e.g., Session 1). Eleven of the 13 explained this by saying that repeated testing with ELISA and Western blot eliminates all false positives. Five of these 13 counselors told the client that false positives had occurred in the 1980s, but no longer today, and 2 said that false positives occur only in foreign countries, such as France, but not in Germany. In addition to these 13 counselors, 3 other counselors initially suggested that false positives do not occur, but became less certain when the client repeated his question and admitted the possibility of false positives (Sessions 2 and 3). Only the three remaining

Table 7.1. *Summary of the Information Provided in 20 AIDS Counseling Sessions*

	100% Certainty	≥99.9%	≥99%	>90%	Range	Best Estimate from the Literature
Sensitivity	5 (of 19)	5	6	3	90–100%	99.99%
Specificity	13 (of 19)	3	3	0	99.7–100%	99.99%
Prevalence	–	–	–	–	0.0075–6%	0.01%
PPV	10 (of 18)	5	1	2	90–100%	50%

Note: Not all the counselors provided numerical estimates. The verbal assertion "absolutely certain" is treated here as equivalent to 100% certain; verbal assertions such as "almost absolutely certain" and "very, very certain" are classified as ≥99%, and assertions such as "very reliable" are classified as ≥90%.

Source: Gigerenzer et al. (1998)

counselors informed the client right away about the existence of false positives. One of these three counselors (Session 4) was the only one who informed the client of the important fact that the proportion of false positives to true positives is higher in heterosexuals, such as the client.

Recall that the currently available estimates indicate that, of heterosexual German men with low-risk behavior who test positive, only 50% actually have HIV. The information provided by the counselors on this was quite different. Half of the counselors (10 out of 18; 2 repeatedly ignored this question) told the client that if he tested positive, it was absolutely certain (100%) that he had HIV (Table 7.1 and Session 1). He was told by five counselors that the probability is 99.9% or higher (e.g., Session 3). Thus, if the client had tested positive and trusted the information provided by these 15 counselors, he might indeed have contemplated suicide, as many people in this situation have done (Stine, 1996).

How did the counselors arrive at this inflated estimate of the predictive value? They seem to have followed two lines of thought. A total of eight counselors confused sensitivity with the PPV (a confusion also reported by Eddy, 1982, and Elstein, 1988), that is, they gave the same figure for sensitivity as for the PPV (e.g., Sessions 2 and 3). For example, three of these eight counselors explained that, apart from the window period, the sensitivity is 100% and therefore the PPV is also 100%. Another five counselors followed a different line of thought: They erroneously assumed that false positives would be eliminated by repeated testing and, consistent with this assumption, concluded that the PPV is 100%. For both groups, the client's question about the PPV must have appeared to repeat a previous question. In fact, more than half of the counselors (11 out of 18) explicitly introduced their answer to this question with a phrase such as "As I have already said . . ." (e.g., Sessions 1–3).

Table 7.1 shows that two counselors provided estimates of the PPV in the correct direction (between 99% and 90%). Only one of these two (Session 4), however, arrived at this estimate by reasoning that the proportion of false positives among all the positives increases when the prevalence decreases. She was also the only one who explained to the client that repeated testing cannot eliminate all possible causes of false positives, such as a positive test reaction to antibodies wrongly identified as HIV antibodies. The second counselor initially asserted that a positive test result means that an HIV infection is "completely certain," but when the client asked what "completely certain" meant, the physician had second thoughts and said that the PPV is "at least in the upper 90s" and "I can't be more exact."

How Was Statistical Information Communicated? Not one of the counselors communicated the information in terms of natural frequencies, the representation that physicians and laypeople can understand best. Except for the prevalence of HIV, all statistical information was communicated to the client in terms of percentages. The four sessions illustrate this. As a consequence, clients probably did not understand the meaning of what was being communicated to them. Further, some of the counselors did not seem to understand the figures they were communicating. This can be inferred from the fact that several counselors gave the client inconsistent information but did not seem to notice this.

Two examples may serve to illustrate the counselors' unawareness of inconsistency. One physician told the client that the prevalence of HIV in men such as the client is 0.1% or slightly higher and that the sensitivity, specificity, and PPV are each 99.9%. To demonstrate that this information is contradictory, we represent it in natural frequencies. Imagine that 1,000 men take an HIV test. One of these men (0.1%) is infected, and he will test positive with practical certainty. Of the remaining uninfected men, one will also test positive (because the specificity is assumed to be 99.9%, which implies a false positive rate of 0.1%). Thus, two men test positive, and one of them is infected. Therefore, the odds of being infected with HIV are 1:1 (50%), not 999:1 (99.9%). (Even if the physician assumed a prevalence of 0.5%, the odds are 5:1 (84%) rather than 999:1. Note how in this case the odds representation paints a rather more dramatic picture than the probability representation based on percentages.)

Next, consider the information the client received in Session 2. For the prevalence (which the counselor did not provide), assume the median estimate of the other counselors, 0.1%. Again, imagine 1,000 men. One has the virus, and he will test positive with practical certainty (the counselor's estimated sensitivity: 99.8%). Of the remaining uninfected men, three will also test positive (the counselor's estimated specificity: 99.7%). Thus, we expect 4 of the 1,000 to test positive and 1 of these 4 to have the virus. So if the test is positive, the probability of being infected is 25% (one in four), and not 99.8% as the counselor told the client.

This study shows, for a representative sample of public AIDS counseling centers in Germany, that counselors were not prepared to explain to a man with low-risk behavior what it meant if he tested positive for HIV. This is not to say that the counselors were generally ignorant; on the contrary, several counselors gave long and sophisticated lectures concerning immunodiagnostic techniques, the nature of proteins, and the pathways of infection. But when it came to explaining to the client the risk of being infected if he tested positive, they uniformly relied on a representational format that did not serve them well and could have been harmful to their clients.

Diagnostic Insight in Physicians (Case Study II)

To see whether diagnostic reasoning improves when statistical information is represented in terms of frequencies, we decided to manipulate the representational format in an experiment with physicians (Gigerenzer, 1996; Hoffrage & Gigerenzer, 1998). We had previously carried out such an experiment with students as participants, with the result that, when information was presented in natural frequencies (rather than in probabilities), the percentage of Bayesian solutions increased from about 16% to 46% (Gigerenzer & Hoffrage, 1995). Yet, it remained an open (and interesting) question whether this result would generalize to experts such as physicians. Medical textbooks typically present information about sensitivity, specificity, and priors in probabilities (as in Figure 7.2, left-hand side). Medical experts may be so "spoiled" by this common practice that they do not appreciate the advantages of representing the relevant information in natural frequencies.

Forty-eight physicians participated in this study (Gigerenzer, 1996; Hoffrage & Gigerenzer, 1998). The physicians were asked to work on four diagnostic problems: inferring colorectal cancer on the basis of a positive hemoccult test, inferring the risk of breast cancer from a positive mammography test, inferring ankylosing spondylitis on the basis of a positive HL antigen B 27 test, and inferring phenylketonuria from a positive Guthrie test. Two versions of each of the four diagnostic problems were presented to the participants. In one version, the relevant information was presented in a probability format; in the other, in a frequency format. (Which problems were in which format and which format was presented first was systematically varied, with the constraint that the first two problems had the same format.) To illustrate, here are the two versions of the colorectal cancer problem:

> To diagnose colorectal cancer, one of the tests that is conducted to detect occult blood in the stool is the hemoccult test. This test may be used for people above a particular age and in routine screening for early detection of colorectal cancer. Imagine that you are screening in a

certain region using the hemoccult test. For symptom-free people over 50 years old who are screened by the hemoccult test in this region, the following information is available:

The Probability Format

The probability that one of these people has colorectal cancer is 0.3%. If one of these people has colorectal cancer, the probability that he or she will have a positive hemoccult test is 50%. If one of these people does not have colorectal cancer, the probability that he or she will still have a positive hemoccult test is 3%. Imagine a person (aged over 50, no symptoms) who has a positive hemoccult test in your screening. What is the probability that this person actually has colorectal cancer? ______ %

The Natural Frequency Format

Thirty out of every 10,000 people have colorectal cancer. Of these 30 people with colorectal cancer, 15 will have a positive hemoccult test. Of the remaining 9,970 people without colorectal cancer, 300 will still have a positive hemoccult test. Imagine a sample of people (aged over 50, no symptoms) who have positive hemoccult tests in your screening. How many of these people do actually have colorectal cancer? ______ out of ______

The physicians received a booklet containing all four problems, two of which presented information in probabilities and two in natural frequencies. The formats and order of the problems were systematically varied among the physicians, with the constraint that the first two problems were in the same representational format. Participants were invited to make notes, calculations, or drawings while working on the problems; these were analyzed later to reconstruct their reasoning. After the physicians had filled out the booklets, we interviewed them about their reasoning strategies. We only coded answers to the problems as being in accord with Bayes's rule when (1) the numerical estimate was within 5 percentage points of the correct one, and (2) the physician's notes, calculations, or drawings, and the interview confirmed that the answer was neither a guess nor the result of another strategy. Next, we describe the impact of the two representational formats on a physician's reasoning. This physician was a rather typical case, and we therefore call him Dr. Average.

A Physician's Diagnostic Reasoning. Dr. Average is 59 years old, director of a university clinic, and a dermatologist by training. He spent 30 minutes on the four problems and another 15 minutes discussing the results with the interviewer. Like many physicians, he became visibly nervous when working on the problems, but only when faced with the probability formats. At first, Dr. Average refused to write notes; later, he agreed to do so, but only on his own piece of paper and not on the questionnaire. He did not let the interviewer see his notes.

Dr. Average's booklet started with the mammography problem in the probability format. He commented: "I never inform my patients about statistical data. I would tell the patient that mammography is not so exact, and would in any case perform a biopsy." He estimated the probability of breast cancer after a positive mammography as 80% + 10% = 90%, that is, he added the sensitivity to the false positive rate (an unusual strategy). Nervously, he remarked: "Oh, what nonsense. I can't do it. You should test my daughter; she studies medicine." Dr. Average was as helpless with the second problem, ankylosing spondylitis, in a probability format. This time he estimated the posterior probability by multiplying the base rate by the sensitivity (a common strategy in statistically naïve students; see Gigerenzer & Hoffrage, 1995).

Then came the first problem presented in a frequency format. Dr. Average's nervousness subsided visibly. Coming up with the Bayesian answer, he remarked with relief: "That's so easy." He also arrived at the Bayesian answer with the fourth problem, which was also presented in a frequency format. Dr. Average's reasoning evidently turned Bayesian when the relevant information was presented in frequencies. This was the case, despite the fact that he did not know Bayes's rule, as he informed us.

Incidentally, Dr. Average was not the only physician who referred in despair to a daughter or son. In one case, the daughter was actually nearby and was also working on the problems. Her father, a 49-year-old private practitioner, worked for about 30 minutes on the four problems and failed on all of them. "Statistics is alien to everyday concerns and of little use for judging individual people," he declared. He derived his numerical estimates from one of two strategies: base rate only or sensitivity only (both strategies are common with statistically naïve students). His 18-year-old daughter solved all four problems by constructing Bayesian trees (as on the right-hand side of Figure 7.2). When she learned about her father's strategies, she glanced at him and said: "Daddy, look, the frequency problem is not hard. Couldn't you do this one either?" For this private practitioner, even frequency formats didn't help. In contrast, a 38-year-old gynecologist faced with the mammography problem in

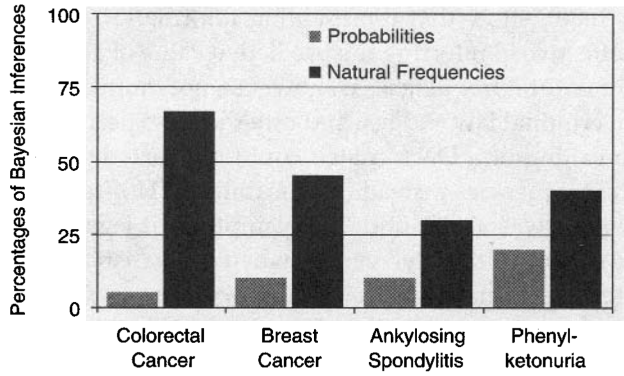

Figure 7.3 The percentage of Bayesian answers is higher when the relevant information is represented as natural frequencies rather than as probabilities.

the frequency format, exclaimed: "A first grader could do this. Wow, if someone couldn't solve this . . .!"

Does Representation Have an Effect on Physicians' Diagnostic Reasoning? On average, the 48 physicians worked on the four problems for half an hour. When the information was presented in a probability format, the physicians reasoned the Bayesian way in only 10% of the cases, averaged across all four problems. When the information was presented in natural frequencies, this figure increased to 46%.

As can be seen in Figure 7.3, the frequency representation led to higher proportions of Bayesian estimates for each of the four problems. In addition, the natural frequencies turned out to be less time-consuming for the participants. The physicians spent about 25% more time on the probability problems, which indicates that they found them more difficult to solve. As the case of Dr. Average illustrates, the physicians often reacted differently – cognitively, emotionally, physiologically – to the probability format and the natural frequency format. The physicians were more often nervous when information was presented in terms of probabilities. When working on probability problems, they made complaints such as: "I simply can't do that. Mathematics is not my forte." However, with natural frequencies, a typical remark was: "Now it's different. It's quite easy to imagine. There's a frequency; that's more visual." In addition, they were less skeptical about the relevance of statistical information to medical diagnosis when it was communicated in frequencies.

The results of this case study, which have recently been replicated with 96 advanced medical students (Hoffrage, Lindsey, Hertwig, &

Gigerenzer, 2000), show that representing information in natural frequencies is effective in inferring the predictive value of a test. The beneficial effect of natural frequencies is, however, not limited to the field of medicine. In criminal law, judges' and other legal experts' understanding of the meaning of a DNA match could similarly be improved by using natural frequencies instead of probabilities (Hoffrage et al., 2000; Koehler, 1996). It was also found, for example, that legal experts were less likely to support a "guilty" verdict when the statistical information was presented in natural frequencies. An important new finding is that natural frequencies can also facilitate reasoning in complex Bayesian situations characterized either by two or more predictors or by predictors and criteria with more than two values (Krauss, Martignon, Hoffrage, & Gigerenzer, in review).

These results have two implications. First, because information in medical textbooks is routinely communicated in terms of probabilities or percentages, medical students as well as physicians ought to be taught how to translate these figures into natural frequencies. Sedlmeier and Gigerenzer (2001, Study 1) designed a computerized tutorial system that teaches people how to do this. People who were taught to translate probabilities into natural frequencies performed twice as well on Bayesian inference problems as people who were taught the standard method of inserting probabilities into Bayes's rule. Even more striking, performance in the group that translated information into natural frequencies remained stable in a 5-week follow-up test (median performance 90% correct), whereas performance in the standard group showed the usual deterioration due to forgetting (15% correct). Kurzenhäuser and Hoffrage (2002) applied this approach to a typical classroom setting and found that twice as many medical students who learned to actively translate probabilities into natural frequencies (as compared to a control group who learned Bayes's rule) were able to deal with probabilities when tested 2 months later.

A second, equally important implication concerns the communication of risks, not only in medical textbooks but also to patients. For instance, before consenting to medical treatment on the basis of a diagnosis, patients should understand the uncertainties involved, such as the risks of actually having the disease. In order to facilitate accurate assessment of risk, physicians should use the most effective means of representation and thus of communication. As we showed, it is not for lack of contenders that more effective representations are not generally available to medical experts and their patients.

Representing Change

Observing Historical Spectra: The Differential Calculus

Isaac Newton (1642–1727) and Gottfried Wilhelm Leibniz (1646–1716), the two eminent figures credited with the intellectual breakthrough of the differential and integral calculus, had a bitter dispute over priority (Hall, 1980). In the work of both Leibniz and Newton, the inception of the calculus was embedded in a body of questions concerned with natural philosophy, metaphysics, and theology (Bertoloni Meli, 1993; see also the Leibniz–Clarke correspondence in Alexander, 1956). Both of them had arrived at the insight that two old problems could be viewed as inverse to each other: finding tangents to curves and finding areas below curves (quadratures). Since antiquity, these had been distinct problems (Boyer, 1949). Although the calculation of tangents and areas had advanced quite far in the century prior to Newton and Leibniz, it was only through their work that these were seen as inverse and that a general and algorithmic method was established. In the centuries following Newton and Leibniz, some of their concepts were recast and abandoned, and new ones, like the limit concept and the function concept, were introduced. The important point is that the institutionalization of the differential calculus began with rivaling proposals and that this plurality of representation has remained a feature of the calculus. To demonstrate this point, we will ask you, the reader, to look at the graph shown in Figure 7.4 through "Leibnizian glasses," "Newtonian glasses," and "Modern glasses."

Let us start with the *Leibnizian glasses*. Leibniz conceived of smooth curves as polygons with infinitely many sides. If you look at the curvilinear line in Figure 7.4 through Leibnizian glasses you see the smooth curve as a chain of short rectilinear line segments. The "links" in this chain may be infinitely small, but the curve will always remain a chain when you are looking through your Leibnizian glasses. The passage from this imagined chain to the quantification of change can be achieved by the following consideration pertaining to the vertices of the (infinite angular) polygon, or, in other words, to the points at which the (infinitely small) links of the chain meet. The ordinates and abscissas corresponding to the vertices of the polygon can be understood as forming number sequences. To approximate the smooth curve, the differences between the successive terms of such sequences were conceptualized as infinitely small. These infinitesimal differences were called *differentials*

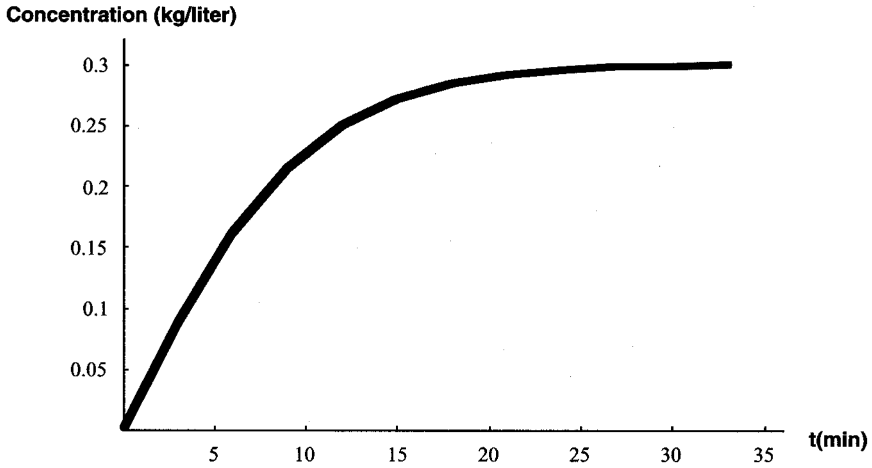

Figure 7.4 Cartesian coordinate system with a graph showing how the salt concentration in the flask changes with time. The line representing the changing concentration can be read in three different ways: a Newtonian, a Leibnizian, or a function-based way.

and denoted by dx or dy, the differentials being assigned to the finite variables x and y, respectively. In other words, the operator d, as in dx, related to a sequence the corresponding difference sequence. Leibniz's operator $\int$, by contrast, related to a sequence the corresponding sum sequence. Thus, Leibniz's insight of the inverse nature of the problem of quadratures and tangents was in the end, or rather in its beginning, based on the inverse operations of summing and finding differences (Bos, 1993).

Now let us try on the *Newtonian glasses*. In Newton's fluxionary calculus, variable quantities, also called *fluents*, were conceived of as changing over time. In Newton's own words (cited from his *Tractatus de quadratura curvarum* of 1704, as translated and reprinted in Struik, 1969, p. 303):

> I consider mathematical quantities in this place not as consisting of very small parts; but as described by a continued motion. Lines are described, and thereby generated not by the apposition of parts, but by the continued motion of points; superficies [surfaces] by the motion of lines; solids by the motion of superficies; angles by rotation of sides; portions of time by a continual flux: and so in other quantities. These geneses really take place in the nature of things, and are daily seen in the motion of bodies.

The velocities or rates of change with respect to time of such variable quantities were called *fluxions*. Newton used *pricked letters* like

$\dot{x}$ to symbolize them. If you look at the graph in Figure 7.4 through Newtonian glasses, you see the curvilinear line as the trace created by a rightward-moving point. Changes in velocity are represented as directional changes on the upward or downward dimension of the rightward-moving point.

The passage from the motion of a geometrical object to the quantification of change was accomplished by considering the change of variable quantities during indefinitely small time intervals, also called *moments* and denoted by the letter o. For instance, Newton determined the momentary change in the area below a curve by adding a moment o to the variable quantity noted on the abscissa and then adding a corresponding term to the variable quantity noted on the ordinate: Consider, for example, the case in which the area underneath a curve is described by the standard equation for a parabola $z = x^2$; adding a moment o to the abscissa corresponds to the expression $z + yo = (x + o)^2$, where z denotes the area, and x and y are the variable quantities denoted on the axes. The computational procedure was then completed by expanding the expression on the right-hand side (applying the binomial theorem), removing the terms without o (because they are equal), dividing by o, and then neglecting terms carrying o as a factor. The result of these computations is $y = 2x$. By this procedure, it was established that the area z underneath the curve $y = 2x$ is x^2. Later, Newton expressed the monetary change of a variable quantity (for example, y) by multiplying its fluxion ($\dot{y}$) by an infinitely small time interval (leading to the expression $\dot{y}o$), but the main idea behind the computational procedure, namely, to consider the change in area by adding very small increments to the variables in an equation, remained the same (see Bos, 1980, pp. 56–59, for a thorough exposition). Newton's insight concerning the inverse nature of the problem of quadratures and tangents was helped by his early work on the binomial theorem and on series expansion in general, and was related to the way in which Newton determined the area underneath a curve.

Lastly, let us switch to the *Modern glasses*. Modern glasses are not part of modern analysis in the same way that Leibnizian and Newtonian glasses had been part of their respective calculi. Geometrical considerations were indispensable for the development, presentation, and justification of Newton's and Leibniz's calculi. The development of modern analysis was accompanied by a distancing from geometrical considerations. The French mathematician Augustin-Louis Cauchy (1789–1857) was a major figure in the reworking of the foundations of the calculus and in its modern formulation. He also introduced the modern notation, in which a prime is used to denote the derivative, for example, $f'(x)$. In Cauchy's definition of limit, and thus of the derivative, no reference was made to geometric figures (Grattan-Guinness, 1980), and his textbook *Cours d'analyse de l'École Royal Polytechnique*, published in 1821, did not contain a single diagram!

However, it is nevertheless possible to construct a visual analogue for the concept of derivative. We ask you to look once again at the graph in Figure 7.4 and to choose a small segment of the curvilinear line. We suggest you imagine that you can move a small open rectangle (rather like a cursor on a computer screen) along the curvilinear line. You could then make it your goal to move the cursor so that the open rectangle frames a segment with particularly strong curvature, but it is up to you. Once you have settled on a segment of the graph, we ask you to imagine that the rectangle and its contents are enlarged so as to fill your entire field of vision, or, in other words, that you are zooming in on the segment. As a result, you see a line with less curvature than the original line segment – you might even see a straight line. Starting with this "new" line, repeat the procedure of choosing a line segment and zooming in on it, and then repeat this procedure as many times as you wish. You will soon get bored because you will be looking at what appears to be the same straight line. With Modern glasses you can thus explore the inner workings of the graph, whereby you will observe that you are approaching a world of straight lines.

To summarize the three different ways of looking at the graph in Figure 7.4: We first looked through Leibnizian glasses and imagined the curvilinear line approximated by a chain with infinitely many infinitely small links; then we looked through Newtonian glasses, imagining the curvilinear line as the trace generated by a moving point; finally, we constructed Modern glasses, zooming in on particular line segments and exploring the inner workings of the graph. The important point for this chapter is that since its inception, the calculus has provided multiple representations of change. This multiplicity is also reflected in the calculus-specific notation. Present-day experts use both Leibnizian notation and Newtonian notation, as well as function-based notation.

Experts' Representation of Change

In this section, we summarize a case study in which calculus experts – a mathematician, a chemist, and a physicist – were asked to solve a problem that requires a differential equation for its exact solution. Next, we introduce the task that these experts were asked to solve, considering first the general model of exponential change.

Task Analysis. What do cell growth, population growth, continuously compounded interest, radioactive decay, cooling of a body, and the decrease in intensity of light when transmitted through a sample have in common? All of these phenomena concern change: an increase or decrease in cell mass, in population size, money, temperature, or light intensity. Furthermore, all of these phenomena have been described, given certain additional assumptions, by a particular mathematical model known as the *law of exponential change*. For example, in an ideal environment, the change in mass of a cell will be proportional to the mass of the cell, at least early on. This relationship can be described by a differential equation (here in Leibnizian notation) of the form:

$$\frac{dx}{dt} = kx, \qquad (3)$$

where x is the mass of the cell, t is time, and k is a constant. This differential equation can be solved to determine the mass of the cell at a particular time t or, for that matter, at any time t. Solving the equation, which requires the operation of integration, leads to the following expression:

$$x = x_0 e^{kt}, \qquad (4)$$

where x_0 denotes the initial mass of the cell. In words, cell mass grows exponentially. The identical equation has been used to model continuously compounded interest, or radioactive decay, or the change in light intensity when light is transmitted through a sample of a certain thickness. (In the last two cases the constant k has a negative value.)

Now consider a slightly more involved situation as represented by, for instance, the cooling of a body in a surrounding medium of constant temperature. In this case, we have to take into account that the body's temperature adjusts to the temperature of the surrounding medium, so that the difference between its temperature and that of the surrounding medium is crucial. (We assume that the temperature of the body does not affect the temperature of the surrounding medium.) *Newton's law of cooling* (here in Newtonian notation) addresses this situation:

$$\dot{T} = -k(T - T_s), \qquad (5)$$

where $\dot{T}$ signifies the change in temperature with time (in Leibnizian notation denoted by dT/dt), k is a constant, and T_s is the temperature of the surrounding medium. In words, the cooling of a body is proportional to the difference between its temperature and that of the surrounding medium. The differential equation describing Newton's law of cooling can also be used to model phenomena other than the cooling and heating of bodies. The expert participants in our third case study worked on a task that can be modeled by an equivalent first-order linear ordinary differential equation.

The following problem, henceforth called the *Flask Problem* (Brenner, 1963), was presented to the expert participants:

> A flask contains 10 liters of water, and to it is being added a salt solution that contains 0.3 kilogram of salt per liter. This salt solution is being poured in at a rate of 2 liters per minute. The solution is being thoroughly mixed and drained off, and the mixture is drained off at the same rate, so that the flask contains 10 liters at all times. How much salt is in the flask after 5 minutes?

This problem can be represented by the following differential equation in Leibnizian notation:

$$\frac{dx}{dt} = 0.2(0.3 - x) = 0.6 - 0.2x \qquad (6)$$

or, in Newtonian notation:

$$\dot{x} = 0.6 - 0.2x. \qquad (7)$$

Solving the equation and taking into account that initially there is no salt in the flask, the answer to the problem is

$$x = 3 - 3e^{-0.2\times 5} \text{ kilograms} \qquad (8)$$

or 1.9 kilograms (rounded to one decimal position).

A notable feature of this problem is that it requires a conceptualization of instantaneous change. Consider, for example, the following "mutilation" of the problem: "A flask contains 10 liters of water, and to it is being added a salt solution that contains 0.3 kilogram of salt per liter. This salt solution is being poured in at a rate of 2 liters per minute. How much salt is in the flask after 5 minutes?" The answer is, of course, 3 kilograms, little more than a multiplication exercise. In this version, the problem still requires one to consider the rate of change of incoming salt, but there is no need (cognitively speaking!) to operate with the concept of instantaneous change; in the "full" version of the Flask Problem, it is necessary to conceptualize instantaneous change and to operate with it. An exciting feature of calculus is that it provides more than one way to do this.

Representational Practices of Differential Calculus

The experimenter met individually with the experts, usually in the expert's office. The participants were allowed to use paper and pencil and a pocket calculator, but no access to reference books was permitted; they were not told that the problem requires calculus for its solution. The participants were asked to think out loud, using instructional materials adapted from Ericsson and Simon (1993). The protocols were taped and transcribed. The protocols in conjunction with the experts' handwritten notes were analyzed into problem-solving episodes. The analyses of three sessions are summarized in the following (for details see Kurz, 1997).

Session 1: A Mathematician. Participant T is a young, highly productive mathematician whose major field is analysis. He is a faculty member in a doctoral-level mathematics department. He worked on the Flask Problem for about 25 minutes; his protocol consisted of 11 episodes.

After reading the problem statement (Episode I) and drawing a schematic picture of the flask with arrows representing inflow and outflow of mixture (Episode II), Participant T assigned variables and briefly pursued an algebraic approach (Episode III). But then he realized that he "should probably use some calculus, in the sense of rates of change" (Episode IV). More specifically, he realized that a "derivative with respect to time" was needed. But he had to admit, somewhat embarrassed, "not seeing how to do this straightforwardly with calculus either." He proceeded by computing the amount of salt that was added after 1 minute (Episode V). He saw that an extrapolation from there to the solution of the problem was not feasible. As a way out of the dilemma, he introduced the concept of *instantaneous rate of change*, which opened possibilities for computation: In his words, "Instead, I wanna try to figure out what's the instantaneous rate of change of, well, what's the saline solution after any given time. So let me go to 30 seconds."

His new strategy was to "refine until nothing," to choose *decreasing* fixed time increments until the increments would become infinitely small. In actual fact, Participant T's computations concentrated on the first and second 30 seconds (Episode VI), and he only considered, hypothetically, to work with 15-second increments (Episode IX). These computations turned out to be rather laborious, involving checking and rechecking of the results (Episodes VII and VIII). He was clearly feeling uneasy with the progress he was making on the problem.

This rather laborious process of "refining" was not unlike perceptual rehearsal in that Participant T repeatedly carried out very similar computations (Episodes V–VIII) that allowed him to "see" a new pattern (see Ippolito & Tweney, 1995; Tweney, 1996). He noticed that "the *rate in* is always the same" (Episode VIII). This inconspicuous insight enabled a crucial next step in the solution, namely, to "Figure out the *rate out*" (Episode X). He then noticed that "it look[ed] like [he was] coming up with the differential equation here." Once represented in this form, it was only a routine task for him to solve the equation. Unfortunately, however, his differential equation was not entirely correct. The value for the rate out was off by one decimal position because he had not taken into account that the incoming salt was dissolved in 10 liters of fluid. For this reason, his final solution could not be interpreted meaningfully and remained unsatisfactory to him, but at this point, frustrated and pressed for time, he was not prepared to "debug" his solution.

Session II: A Chemist. Participant U is a midcareer physical chemist and a faculty member in a doctoral-level chemistry department. She is very active in research and has published many papers in her field. Participant U spent approximately 40 minutes working on the problem; her protocol consisted of 12 episodes.

Participant U spent considerable time (about 10 minutes) reading (Episode I) and rereading the problem (Episode II). At the end of her second reading she singled out "the critical sentence here," namely, that "the solution is being thoroughly mixed and drained off" (Episode III). From there she reasoned that "the concentration would be increasing over a period of a few minutes, and at some point you'd reach a steady state where you were putting the same amount of salt in as was going out" (Episode IV). This understanding led to a graphical representation (similar to the graph in Figure 7.4). She drew a Cartesian coordinate system with time on the abscissa and salt concentration on the ordinate. Then she constructed a "graph in time" (Episode V) by going to the "1-minute" location in the coordinate system and marking the respective value for the salt concentration, next to the "2-minute" location and again marking the respective value for the salt concentration, and so forth. For the first minute, she explicitly assumed that no salt was leaving the flask; for the 2-minute location she made sure, after adding the same amount of salt than for the first minute that this time she marked a point that was "a little low," because salt loss had to be taken into account. She proceeded to the "3-minute" location, adding twice the amount of the first minute, and again making sure that the value on the ordinate was "low, we'd be even lower." The essential feature of the resulting graph was its asymptotic nature, in her words, that "it's gonna be coming up like this and then it's just gonna be a straight line for the rest of the time" (Episode VI). The change in salt concentration with time had been transformed into the continuous motion of a point, generating a line. In this sense, her graph was a dynamic representation, that is, a representation in which time is necessarily represented in an analog fashion (Freyd, 1987).

Participant U also meant to use her graph to infer a solution to the Flask Problem, but then she noticed that the scale on the ordinate was not right. In fact, the scale was off by one decimal position. She did not read off a solution from the graph but instead assumed that the concentration after 5 minutes would be the steady-state concentration of 0.3 kilogram – the concentration of the incoming salt solution. This assumption gave her 3.0 kilograms as the amount of salt after 5 minutes. The experimenter, somewhat worried that Participant U would quit at this stage, asked whether she could formulate an equation. Immediately she wrote dc/dt (c for concentration, t for time), the "change in the concentration," to represent the left-hand side of an equation (Episode VII). For the right-hand side she reasoned that "first, the concentration is zero," so "the intercept is zero." She added a term denoting "the increase in the concentration," which was 0.6, with the units "kilograms per liter per minute." The "concentration going out," being the "minus part," she approached by making a "linear assumption" (Episode VIII). But she quickly realized that this assumption led to the anomaly of "not getting a steady state." Thus, it had to be a nonlinear function. But what kind of nonlinear "functional form"? This was a difficult question that led her to recapitulate what she knew about the physical process, namely, that "it's increasing at a constant rate" and that "the concentration is increasing linearly," and she was certain "that it's not decreasing at a constant rate," and also that a constant volume of the perfectly mixed solution was pouring out (Episode IX). After silently rereading parts of the problem, she emphasized that she was thinking about the process "in terms of a continuous thing," and that the approximation method she was going to propose next would be a "quickie" way to arrive at a solution (Episode X).

Participant U's approximation method (Episode XI) matched the way she had previously constructed her graph. In the first minute, 0.6 kilogram of salt was added to the flask, and she assumed that no salt was deleted from the flask during this time. In the second minute, another 0.6 kilogram was added and one-fifth of the amount of salt in the flask after 1 minute was subtracted, because the salt was dissolved in 10 liters, 2 of which were withdrawn during the second minute. With the third minute, another 0.6 kilogram of salt was added to the amount in the flask and one-fifth of the amount of salt in the flask after 2 minutes was subtracted. She carried this procedure through to 5 minutes and then announced her solution as "2.14 kilograms in 10 liters; that's approximate!" The experimenter asked what she would do to improve her approximation. She answered promptly (Episode XII): "Well, you have to take smaller time intervals." Finally, the experimenter asked for her best guess of the precise solution. She answered: "Oh, 1.9 kilograms or something at 5 minutes." Certainly, an excellent guess. In a way, she had achieved, what she had proposed earlier (Episode X), namely, that after her solution by approximation she would "try to work back so that [she'd] have an instantaneous picture of what was going on."

Session III: A Physicist. Participant S is a theoretical physicist internationally known for contributions to his field. He teaches undergraduate and graduate physics courses in a masters'-level physics department. Participant S spent about 50 minutes working on the Flask Problem; his protocol consisted of 18 episodes.

After reading the problem statement (Episode I), Participant S determined that his task was to formulate a model, in his words, "to put all this together in some formulas or something and see these relationships" (Episode II). He restated the problem in his words and then computed a "guess," a numerical solution based on simplifying assumptions (Episode III and IV). For his guess, he assumed that salt solution was added to the flask at one instant and deleted at another. He announced 1.5 kilograms to be his answer to the problem, adding that it "may not be right." Fearing that, with this answer, Participant S might end his problem solving, the experimenter asked, "Can you come up with an equation?" "A good question," Participant S agreed, because equations had been "implicit" in what he had been thinking, but now the challenge was to "find what they are." It was clear to him that it would need to be "sort of a rate equation" (Episode V). He started to wonder whether he had missed something before and therefore thought it best to "read this again" (Episode VI).

After rereading the problem, he assigned x to "the amount of salt in the tank" (Episode VII). The left-hand side of his *rate equation* was "the time derivative of x" and was noted in Newtonian notation (see Equation 9). The right-hand side had to be "the rate at which it's added minus the amount that is leaving" (Episode VIII). The "rate at which it's added" was 0.6 kg/min and was followed by a minus sign. He immediately recast the rate of incoming salt as 0.3(2), omitting the units of measurement. Then he proceeded to the part following the minus sign, which had to be "also a function of time." He knew that "the amount of fluid that is flowing out is 2." But how did the amount of salt that was leaving the flask depend upon the concentration of salt solution in the flask? The amount of salt in the tank "is always gonna be x over 10." He had written out the complete differential equation:

$$\dot{x} = 0.3(2) - \frac{x(2)}{10}. \tag{9}$$

After a lengthy pause (of 12 seconds), he came to the conclusion "that this might be the right idea, really," cause this says that the rate at which the amount of salt changes depends upon how fast you add it" (Episode IX). He checked the units of measurement (Episode X) and was just delighted to find that the formulated equation had "the right units, this has the right units!"

He anticipated that solving this equation would take some effort on his part. He restated the equation in Leibnizian notation, which is preferable for solving differential equations. (When he was asked later about this switch in notation, it turned out that he had not been aware of it.) But before actually beginning to solve the equation, he wanted to "see whether [he] like[d] the way this [was] going" (Episode XII), whether this mathematical model matched his process understanding. He determined that "in the extreme future you would reach an equilibrium situation where all of the original water had been replaced and therefore the concentration inside the tank would be 0.3 kilogram per liter." He thus was able to determine the amount of salt in the "extreme future," but the problem asked for "the answer in the middle." He thought that he would find out "the answer to this question by solving this equation," and therefore "it's worth doing" (Episode XII).

Solving his rate equation (Episode XIII; separating variables, then using a substitution procedure, in which he differentiated with respect to a dummy variable y, and then integrating the resulting expression), he arrived at an intermediary result which he evaluated at $t = 0$ and for $t \to \infty$ (Episode XIIV). He found that the equation did not exhibit the right behavior at $t = 0$ when no salt was supposed to be in the flask. But nevertheless, he detected "elements of truth here" because he "saw" that "at long times" the model would reach the appropriate equilibrium. He realized that he had not properly integrated the equation (in his substitution procedure he had used the operations of differentiation and of integration, which made a second integration necessary to solve the differential equation). He integrated, evaluating a definite integral, and computed the general solution (Episode XV). He evaluated the resulting equation at $t = 0$ and for $t \to \infty$ and found what he "thought ought to happen" (Episode XVI). Finally, he substituted $t = 5$ to determine the amount of salt after 5 minutes (Episode XVII) and then used a hand calculator to determine the numerical solution, which was 1.896 kilograms (Episode XVIII).

How Did the Calculus Experts Represent Change?

The solutions worked out by these three experts differed remarkably in many respects. Here the focus is on their use of the differential calculus. In a nutshell, the mathematician's representational use of the calculus was in many respects Leibnizian, the chemist's Newtonian, and the physicist's born out of a genuine modeling approach. Specifically, the mathematician's solution was based on the choice of fixed increments. In the limiting case these decreasing fixed increments are Leibniz's differentials. Limit taking in Leibniz's calculus was global (Bos, 1993, p. 87). With respect to a smooth curve, this global limit taking meant that the curve remained composed of the sides of a polygon even after extrapolation to the infinite case. By contrast, the derivative defines a local limit (see the earlier discussion of the Leibnizian glasses in contrast to the Modern glasses). Participant T knew that a "derivative with respect to time" was necessary, but he did not know how to model the problem using this concept. As a way out of this dilemma, he worked with decreasing fixed increments. In a sense, then, Participant T "approximated" the concept of derivative with his computations rather than the numerical solution of the problem. But even if Participant T's plan to "refine until nothing" was not Leibnizian in intent, the realization in terms of fixed increments was.

The chemist's solution was based on the transformation of the change in salt into the continuous motion of a point, creating a "graph in time." Her approximation method finally paralleled her construction of this graph: First, she determined how much was added; then she made sure that she was "a little low," that she subtracted about the right amount. Next, she extrapolated the appropriate "motion," resulting in her "graph in time." The transformation of change into the motion of a geometrical object was central to Newton's fluxionary calculus (see the previous discussion of the Newtonian glasses). Newton conceived of mathematical quantities as motion of geometrical objects (see the previous quote from the *Tractatus de quadratura curvarum*). Similarly, Participant U provided a successful solution because she was able to utilize a dynamic representation that enabled her to work out an approximation procedure that led to a numerical solution of the problem.

Finally, the physicist engaged in an ongoing process of checking the match between his mathematical model and his physical process understanding, both being constructed simultaneously. In order to consolidate this match, Participant S made the process observable and manipulable. This simulation of the physical process observed a process in time, from "no salt in there to start with" to an "extreme future" in which "you would reach an equilibrium" (Episodes XIIV and XVI). Description was observation in this case; this is, for instance, also parallel to what Nersessian (1992) concluded about thought experiments. And in this case, observation was inextricably coupled with manipulation – a unity that also has been emphasized for experimentation, for instance, in relation to Michael Faraday's experimental investigations (see Gooding, 1992; Tweney, 1992). This unity of observation and manipulation occurred at the interface of Participant S's understanding of the physical process and of his mathematical model. Although it could be argued that the differential equation is the physical process model, the identification of the understanding of the physical process in terms of the mathematical model was the final stage of Participant S's solution process; it was in fact his primary achievement.

This case study shows that there is variability in the representational use of the calculus. Moreover, this variability becomes meaningful when related to the historical development of the differential calculus. In its historical development this variability was also related to differences in the understanding of natural phenomena. Corresponding to their different representations of change, Newton and Leibniz, for instance, also had different notions of accelerated motion (Bertoloni Meli, 1993, pp. 74–91). For Leibniz, accelerated motion was a series of infinitesimal rectilinear motions interrupted by impulses; for Newton, it was a continuous curve where force acts continually. The experts in this case study showed great competence in choosing and developing a representation that was meaningful to them. The plurality of representations provided by the calculus is a feature that experts may use to further their understanding.

Representation Matters

In this chapter, we have argued (1) that the calculi of uncertainty and change provide multiple representations, which bridge the past and present, (2) that the choice of representation is already part of the solution of the problem, and (3) that learning to choose an appropriate representation can help experts to understand uncertainty and change, and to communicate successfully with their clients and students on such topics as how to assess risk. As the studies with AIDS counselors and physicians dramatically demonstrated, the training of experts does not always include learning to choose a suitable representation.

The importance of representations has been emphasized repeatedly, from cognitive science (e.g., Marr, 1982) to physics (e.g. Feynman, 1967). Otto Neurath, who cautioned scientists against wasting their time with history, was in fact a pioneer in designing external representations to help ordinary citizens understand statistical information. Neurath (1939) successfully developed the visual statistical language ISOTYPE – so successfully that in the early 1930s the Soviet government invited him to train specialists to teach the Soviet people ISOTYPE. Unfortunately, he was never paid for his efforts and got into serious financial difficulties (Hegselmann, 1979).

The tools that experts use to quantify risk and change support variability in representational practice. This variability can be beneficial, as we have shown for the case of risk assessment in medical diagnosis, where most experts and laypeople are dramatically helped by representations using natural frequencies. Competence thus can mean knowing how to re-represent a problem so that reasoning is facilitated. Becoming competent in this fashion requires us to acknowledge the possibility of a plurality of alternative representations, in fact, a hallmark of mathematical thinking in general – no matter by whom it is carried out.

In modern terms, the lesson from Neurath, Leibniz, and other students of representation is: Minds do not reason from information, but from representations. Even mathematically equivalent representations can make a difference to the kind of insight experts gain. The power and the indispensable nature of representations was with us in the past, affects us in the present, and will continue to do so in the future.

References

Alexander, H. G. (Ed.). (1956). *The Leibniz–Clarke correspondence*. Manchester, UK: Manchester University Press.

Bertoloni Meli, D. (1993). *Equivalence and priority: Newton versus Leibniz*. Oxford: Clarendon Press.

Bos, H. J. M. (1980). Newton, Leibniz and the Leibnizian tradition. In I. Grattan-Guinness (Ed.), *From the calculus to set theory, 1630–1910* (pp. 49–93). London: Duckworth.

Bos, H. J. M. (1993). *Lectures in the history of mathematics*. Providence, RI: American Mathematical Society.

Boyer, C. B. (1949). *The history of the calculus and its conceptual development*. New York: Dover.

Brenner, J. L. (1963). *Problems in differential equations*. San Francisco: W. H. Freeman.

Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299*, 999–1001.

Cournot, A. A. (1843/1975). *Exposition de la théorie des chances et des probabilités*. Paris: J. Vrin.

Damerow, P. (1995). Abstraction and representation: Essays on the cultural evolution of thinking. In R. S. Cohen, Jürgen Renn, & Kostas Gavroglu (Eds.), *Boston studies in the philosophy of science* (Vol. 175, pp. 173–298). Dordrecht: Kluwer Academic.

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

Deutscher Bundestag (Ed.) (1990). *AIDS: Fakten und Konsequenzen*. Final report of the Enquete Committee of the 11th German Bundestag, 13/90. Bonn: Bonner Universitäts Buchdruckerei.

Dilson, J. (1968). *The abacus: A pocket computer*. New York: St. Martin's Press.

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* (pp. 249–267). Cambridge: Cambridge University Press.

Elstein, A. S. (1988). Cognitive processes in clinical inference and decision making. In D. C. Turk & P. Salovey (Eds.), *Reasoning, inference and judgment in clinical psychology* (pp. 17–50). New York: Free Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Feynman, R. (1967). *The character of physical law.* Cambridge, MA: MIT Press.
Freyd, J. J. (1987). Dynamic mental representation. *Psychological Review, 94,* 427–438.
Gigerenzer, G. (1996). The psychology of good judgment: Frequency formats and simple algorithms. *Journal of Medical Decision Making, 16,* 273–280.
Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704.
Gigerenzer, G., Hoffrage, U., & Ebert, A. (1998). AIDS counseling for low-risk clients. *AIDS Care, 10,* 197–211.
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: Cambridge University Press.
Gooding, D. (1992). Putting agency back into experiment. In A. Pickering (Ed.), *Science as practice and culture* (pp. 65–112). Chicago: University of Chicago Press.
Grattan-Guinness, I. (1980). The emergence of mathematical analysis and its foundational progress, 1780–1880. In I. Grattan-Guinness (Ed.), *From the calculus to set theory, 1630–1910* (pp. 94–148). London: Duckworth.
Hacking, I. (1975). *The emergence of probability.* Cambridge: Cambridge University Press.
Hall, A. R. (1980). *Philosophers at war: The quarrel between Newton and Leibniz.* Cambridge: Cambridge University Press.
Hegselmann, R. (Ed.). (1979). *Wissenschaftliche Weltauffassung, Sozialismus und Logischer Empirismus.* Frankfurt am M: Suhrkamp.
Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine, 73,* 538–540.
Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290,* 2261–2262.
Ippolito, M. F., & Tweney, R. D. (1995). The inception of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 433–462). Cambridge, MA: MIT Press.
Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences, 19,* 1–53.
Krauss, S., Martignon, L., Hoffrage, U., & Gigerenzer, G. (in review). *Bayesian reasoning and natural frequencies: A generalization to complex situations.*
Krüger, L., Daston, L., & Heidelberger, M. (Eds.). (1987). *The probabilistic revolution, Vol. 1: Ideas in history.* Cambridge, MA: MIT Press.
Krüger, L., Gigerenzer, G., & Morgan, M. S. (Eds.). (1987). *The probabilistic revolution, Vol. 2: Ideas in the sciences.* Cambridge, MA: MIT Press.
Kurz, E. M. (1997). *Representational practices of differential calculus: A historical-cognitive approach.* Ph.D. dissertation, Bowling Green State University.
Kurzenhäuser, S., & Hoffrage, U. (2002). Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher, 24,* 531–536.
Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman.
Moon, P. (1971). *The abacus: Its history; its design, its possibilities in the modern world.* New York: Gordon Breach.

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Minnesota studies in the philosophy of science: Vol. 15. Cognitive models of science* (pp. 3–44). Minneapolis: University of Minnesota Press.

Neurath, O. (1930/1931). Wege der wissenschaftlichen Weltauffassung. *Erkenntnis, 1*, 106–125.

Neurath, O. (1939). *Modern man in the making*. New York: Knopf.

Norman, D. A. (1993). *Things that make us smart*. Reading, MA: Addison-Wesley.

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

Quetelet, L. A. J. (1835). *Sur l'homme et le développment de ses facultés, ou essai de physique sociale*. Paris: Bachelier. [*A treatise on man and the development of his faculties*. Translated by R. Knox (1842), Facsimile edition of translation 1969. Introduction by S. Diamond. Gainesville, FL: Scholars' Facsimiles and Reprints.]

Shafer, G. (1989, November 20). *The unity and diversity of probability*. Inaugural lecture University of Kansas.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252–266.

Shapiro, B. J. (1983). *Probability and certainty in seventeenth-century England*. Princeton, NJ: Princeton University Press.

Stine, G. L. (1996). *Acquired immune deficiency syndrome. Biological, medical, social and legal issues*. Englewood Cliffs, NJ: Prentice Hall.

Struik, D. J. (Ed.). (1969). *A source book in mathematics, 1200–1800*. Cambridge, MA: Harvard University Press.

Tweney, R. D. (1992). Stopping time: Faraday and the scientific creation of perceptual order. *Physis: Revista Internazionale di Storia della Scienzia, 29*, 149–164.

Tweney, R. D. (1996). Presymbolic processes in scientific creativity. *Creativity Research Journal, 9*, 163–172.

Zhang, J., & Norman, D. A. (1995). The representation of numbers. *Cognition, 57*, 271–295.