Introduction

Even the birth of probability theory, which is usually dated from a correspondence between Blaise Pascal and Pierre Fermat in 1654, was connected with a probabilistic uneasiness: Both mathematicians worked on two problems on behalf of the Chevalier de Méré and – although they could solve them – de Méré was unhappy: Not, because the solutions were too sophisticated, but because they contradicted his intuition.[1] Abraham de Moivre summarized the initial difficulties in the taming of chance: "Some problems, that bare posed by chance, at first sight seem very easy; one believes they can be solved with a bit of common sense. Unfortunately this evaluation too often proves to be false and mistakes based thereupon are not rare." (see Krämer, 1996, p. 159).

One century later – during the enlightenment – mathematicians and philosophers believed that these teething troubles had been solved. They were convinced that *formal probability theory* and *human probabilistic reasoning* are just two sides of the same coin. For instance, Pierre Simon de Laplace identified human thinking not as a victim but rather as a "killer" of probabilistic paradoxa. In his *Essai philosophique sur les Probabilités* from 1814 he wrote: "The mind is exposed to fallacies just as the visual sense is. Just like the sense of touch corrects the illusion of the latter, thinking and calculating correct the illusions of the former." Deliberations of this kind prompted him to assume a close match between human thinking and probability theory: "The theory of probability is at bottom nothing more than good sense reduced to a calculus which evaluates that which good minds know by a sort of instinct, without being able to explain how with precision" (Laplace, 1814/1951, p. 196). In 1736, Jacob Bernoulli expressed a similar mental attitude in a letter to Wilhelm Gottfried Leibniz. He speculated that the law of large numbers is a rule that "even the stupidest man knows by some instinct of nature per se and by no previous instruction." (see Gigerenzer et al., 1989, p. 29). The available mathematical tools, in particular the theorem of Bayes and

[1] One of the questions was how many times one has to toss two dice in order to get a double six, and the other question was how to split up the stakes fairly after interrupting a card game (for details see Barth and Haller, 1996, p. 72).

Bernoulli's law of large numbers, were seen as descriptions of actual human judgment (Daston, 1988).

This enlightened view of human cognition lasted until the middle of the 19th century. While Piaget and Inhelder were still claiming in 1956 that "deductive reasoning is the propositional calculus itself", from the late 1960s this view shifted. Modern experimental psychology suggests that humans' skills of statistical thinking are not very distinctive. In the past few decades a flurry of work documented the ways in which actual human reasoning differs from the probabilistic norm. This "heuristics and biases" view on human cognition is summarized by Kahneman, Slovic and Tversky (1982) in the book *Judgment under uncertainty. Heuristics and biases*. Deviations from the probabilistic ideal were regarded as proof that unaided human reasoning is riddled with fallacies. Kahneman and Tversky formulated their credo in 1973: "In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors." This paradigm dominated the research on probabilistic thinking until the nineties. Piattelli-Palmarini, for instance, reinforced this view in 1991: "We are a species that is uniformly probability-blind, from the humble janitor to the Surgeon General [...]. We should not wait until Tversky and Kahneman receive a Nobel prize for economics. Our self-deliberation from cognitive illusions ought to start even sooner." And Stephen Gould summarized in 1992: "Tversky and Kahneman argue, correctly, I think, that our minds are not built (for whatever reason) to work by the rules of probability."

Yet, the story of the evaluation of humans probabilistic abilities turned into something of a ping-pong match. A new turnaround toward an "enlightened view" of human cognition is exemplified by the work of Gigerenzer and Hoffrage (1995). Both psychologists took a closer look at the *representation of uncertainty* in the tasks used in Kahneman and Tversky's "heuristic and biases" program. They realized that the uncertainty communicated in these tasks was expressed in terms of probabilities or percentages. As we have already seen in the preface, understanding percentages is often difficult *per se*, even when no complicated calculations are required. Gigerenzer and Hoffrage (1995) questioned whether it is possible to infer from the human inability to solve such *probability* tasks to a fundamental lack of mental algorithms for *judgment under uncertainty*. They affirmed that *mental algorithms need information* and *information needs representation*.

Is there a representation of uncertainty that is adjusted to human thinking? Surely probabilities and percentages cannot be observed directly in nature and are rarely processed in "natural" human thinking. What other possibilities are there to represent uncertainty? We can represent probabilistic information in the following *numerical* representations:[2]

Representations of probabilistic information	Example
percentages	40%
decimal numbers	0,4
fractions	$\frac{4}{10}$
absolute frequencies	4 out of 10
odds ratio	4 : 6

Table 0.1: Numerical representations of uncertainty

Which of these representations is best adjusted to human thinking? According to theories about memory and attention, absolute frequencies are one of the categories of information that is registered automatically, i.e., without conscious intention and without interfering with other cognitive processes (*automatic frequency processing*, Hasher & Zacks, 1984). This suggests that participants should be presented with tasks in which uncertainty is expressed in terms of absolute frequencies to ascertain whether this form of representation improves their performance.

The present debate on the role of representation of information to improve human insight (e.g., Gigerenzer and Hoffrage, 1999), we extend in chapters 1 and 2. This debate focuses on one of the most confusing and most controversial formulas of probability theory: Bayes' rule. "Bayesian updating" – i.e., revising the probability of a hypothesis (H) when new data (D) arise – occurs in real-life reasoning in a variety of situations. Of

2 Non-numerical representations are, for instance, *verbal expressions* – such as "pretty sure" or "rather unlikely". *Visual representations* of uncertainty – such as tree diagrams – will be introduced in chapters 1 and 2.

great relevance are especially expert judgments. For example: A judge has to change her belief about the guilt of a suspect (hypothesis H) when new evidence (data D) is put forward. A physician has to change her belief about the state of her patient (hypothesis H) when provided with new test outcomes (data D). In 1995, Gigerenzer and Hoffrage were able to show empirically that replacing probabilities in such Bayesian reasoning tasks by *natural frequencies* – which are a special kind of absolute frequencies – can foster participants' insight dramatically. This is the starting point for the dissertation.

In Chapter 1 the beneficial effect of natural frequencies will be introduced and discussed in detail. The aim of the first chapter is to extend Gigerenzer and Hoffrage's (1995) natural frequency approach to *complex-structured* Bayesian situations: All research concerning natural frequencies so far has only referred to situations in which one binary cue and one binary criterion is provided. For instance, this situation occurs if a physician has to judge the state of illness (binary criterion with the values "ill" and "healthy") based on one medical test (binary cue with the values "test positive" or "test negative"). Real-life decisions often require dealing with more complex situations, such as cues or criteria with more than one value, or solving tasks with more than one cue: For instance, a physician might have to consider the outcomes of two medical tests (two cues), or different diseases might be possible (non-binary criterion). There has been scepticism in the literature regarding whether the natural frequency concept can be extended to such complex situations (Massaro, 1998). In this chapter, we provide empirical evidence that, even in these situations, communicating the statistical information in terms of natural frequencies is both possible and beneficial. The generalization of the natural frequency approach also turned out to be helpful when addressing some current critiques regarding this approach. Chapter 1 is closed by a definition of natural frequencies, which takes complex situations into account, and by discussing the limits of cue-integration.

In Chapter 2 we investigate possibilities of de-constructing "Bayesian brain teasers". As a touchstone for this claim we chose one of the most notorious brain teasers regarding probability theory, namely the *Monty Hall Problem* (or *Three Door Problem*). Apart from the natural frequency concept already discussed, we implemented other psychological manipulations into problem's wording by making use of the following psychological elements: *perspective change*, *mental models* and *less-is-more effect*. By providing new ways of representing the problem, we were able to increase participants' performance to levels well above those found by previous researchers. Furthermore, constructing intuitive wordings of the Monty Hall problem as well as analyzing participants' protocols afterwards reveals synergistic connections between the implemented elements, which in cognitive psychology usually are handled separately.

In Chapter 3, we address *scientific statistical thinking*. Today, significance testing is the widest-spread method for evaluating hyotheses in the social sciences. Interestingly, when testing hypotheses, people tend to interpret the result of significance tests incorrectly and yet in a Bayesian way.

Why is a sound understanding of the *meaning* of significance tests important? The experimenter is often not only responsible for conducting the experiment, she also is responsible for communicating the results of the experiment. This requires real insight into the underlying meaning of such a test result, since from this the following answers must be derived: What has to be communicated to the parties interested in the study? What consequences can be drawn? What does this significant test result actually mean, and, what *cannot* be inferred? In the third chapter, we experimentally demonstrate that the belief in common fallacies about the meaning of a significant test result were also held by many methodology instructors who teach statistics to psychology students. Indeed, we found that – albeit not the statistics professors tested – most methodology instructors shared the misconceptions of their students. We suggest a pedagogical approach to overcome these misconceptions by contrasting hypothesis testing with Bayes' rule. If one wants to prevent students' belief that a significant test result says something about the probability of hypotheses, one should explicate the approach that *actually* can deliver such probabilities.

In the General Discussion, the dissertation concludes with a summary of results and suggestions for future work.

To summarize: The basic hypothesis underlying this dissertation is that statistical information can be represented and taught in ways leading to a robust and a deep understanding of fundamental concepts of probability theory. The main ingredient of the problems chosen is the crucial concept of conditional probabilities.