**Fachbereich Erziehungswissenschaften und Psychologie**

**der Freien Universität Berlin**

# Representations Facilitate Bayesian Reasoning – Computational Facilitation and Ecological Design Revisited

Dissertation

zur Erlangung des akademischen Grades

Doktor der Philosophie (Dr. phil.)

Vorgelegt von

Dipl.-Psych.

Jan Multmeier

Berlin, 2012

| | |
|---|---|
| Datum der Disputation: | 4. Juni 2012 |
| Erstgutachter: | Prof. Dr. Gerd Gigerenzer<br>Max-Planck-Institut für Bildungsforschung<br>Berlin |
| Zweitgutachterin: | Prof. Dr. Katja Liebal<br>Freie Universität Berlin |

# Acknowledgments

This dissertation is the result of three years of research carried out at the Harding Center for Risk Literacy and the Center for Adaptive Behavior and Cognition (ABC) of the Max Planck Institute for Human Development. Along this journey, many interactions with colleagues and friends have left me inspired, reflective, or simply amazed.

First, I wholeheartedly thank Odette Wegwarth whose determination and faith brought me to the Harding Center and ABC in the first place. Odette has shown remarkable patience when I was stuck and provided her full intellectual, but also emotional support in all stages of my work. I could not have wished for a better supervisor.

Second, I thank Gerd Gigerenzer for creating the most stimulating scientific environment I have experienced, and for giving me the opportunity to be a part of it. His many insightful and constructive comments improved the quality of my work, and ultimately made me a better scientist. He taught me to think, read, and write all new.

I also thank David Harding, whose generous gift to Gerd Gigerenzer marked the beginning of the Harding Center for Risk Literacy and payed for my rent.

Regarding the work presented in chapter II, I am indebted to Kirsten Volz, whose comments sparked the idea, and Laura Martignon for the many valuable discussions. Sebastian Wagner helped prepare beautiful icon arrays which have attracted a lot of attention at conferences and informal talks. I thank Fr. Mojem and Fr. Willeke for allowing me to conduct the experiment in their schools and their dedication in convincing teachers and parents about the value of this research. Finally, the study could not have been conducted without the consent of the parents and the curiosity of their children to learn a different type of math.

I thank Maria Harde who assigned me the topic of developmental impairments in number processing in a required course of educational psychology at the University of Bielefeld in 2006. Unknowingly, she planted the seed of what would eventually develop into the research presented in chapter III. Michael von Aster provided insightful ideas about the study design. I thank Solveig Wittfoth for her help with preparing the study materials, and Grit Rudinger for her outstanding support in recruiting the subjects and collecting the data. Again, this study could not have been realized without the support of educational centers for children with developmental impairments. Of all those who generously provided their time and infrastructure, I especially thank Hr. Wieneke from *Zentrum zur Therapie der Rechenschwäche* and Hr. Kind from *Lernpraxis Berlin* for their personal involvement in the project, as well as all children who participated in this study and their parents for their commitment.

I thank Dominique Schmidt for programming the task used in chapter IV, and Gregor Caregnato for collecting the data incredibly fast. I warmly thank Wolfgang Gaissmaier for countless helpful comments and discussions about the data analysis and previous versions of the chapter.

Finally, I am thankful for the many helpful and humorous comments by Kibby McMahon on earlier versions of this work and her help with the English wording.

More personally, I am grateful to have parents and friends who gave me support in the rough patches of this dissertation, and celebrated the joyful moments with me. Thank you all.

# Table of Contents

# I – Introduction

## Are Humans Bayesian?

Lacking divine omniscience, humans are left in the mist of uncertainty about the world, relying solely on their ability to use the available information to navigate their environment, predict the future, and ultimately make advantageous decisions for themselves and their social group.

Before leaving our house to go to work in the morning, we might wonder if we should take an umbrella with us or if it is safe to leave it home. We might reckon that it is November and rain is rather frequent at this time of the year. And we might have a look at the sky and see dark clouds approaching. Finally, we might take both into account – our prior assumptions about the likelihood of rain on any given day in November, and our observation that dark clouds are approaching, signaling likely precipitation. Judging the likelihood of rain that way engages us in Bayesian reasoning, i.e. the process of integrating prior beliefs about the likelihood of a hypothesis (it will rain soon) with new data to form an updated belief about the hypothesis. More formally, it describes how the prior probability of a hypothesis is integrated with new information, which may strengthen or weaken the hypothesis. Bayesian reasoning is not only involved when predicting the weather, but pervade all aspects of life. For example, John Staddon has argued that the learning mechanisms behind habituation, sensitization, classical conditioning and operant conditioning can be formally described as Bayesian reasoning models (Staddon, 1988).

Let us consider the classical example from the domain of medicine, of a doctor who has to judge whether a patient with a positive mammography result actually has breast cancer (adapted from Eddy, 1982). There are two mutually exclusive and exhaustive hypotheses: the patient may have breast cancer (the hypothesis, H), or she may not have breast cancer (¬H). The doctor also knows the base-rate of breast cancer for women with similar characteristics as the patient, i.e. the *prior* probability of breast cancer for any given woman (p(H)). Also, there is new data to support the hypothesis that the patient has breast cancer, the positive mammography result (D). Let us assume the doctor has the following information available:

1. The probability of breast cancer is 1% for a woman age 40 who participates in routine screening (p(H) = 0.01)
2. If a woman has breast cancer, the probability is 80% that she will get a positive mammography (p(D|H) = 0.8)
3. If a woman does not have breast cancer, the probability is 9.69% that she will also get a positive mammography (p(D|¬H) = 0.0969)

What is the probability that she actually has breast cancer, i.e. the *posterior* probability p(H|D), considering the prior probability of breast cancer and the positive mammography result? The answer is given by applying Bayes rule:

$$p(D|H) = \frac{p(H)\times p(D|\mathrm{H})}{p(H)\times p(D|\mathrm{H})+\mathrm{p}(\neg\mathrm{H})\times \mathrm{p}(\mathrm{D}|\neg\mathrm{H})} = \frac{0.01\times 0.8}{0.01\times 0.8+0.99\times 0.0969} = 0.078 \quad (1)$$

Combining the information according to Bayes rule, the probability that a woman who tests positively in a routine mammography screening actually has breast cancer is 7.8%.

Research pioneered by Ward Edwards (1968; Phillips & Edwards, 1966) investigated if humans actually integrate information about prior probabilities and evidence according to Bayes rule. Based on his experiments, he concluded that probability estimates were generally proportionate to Bayesian posterior probabilities. Judgments of posterior probabilities changed with the amount of evidence for or against a hypothesis, but he also noted that estimates were generally "conservative", in the sense that people used new information too little. Applied to the mammography example, conservative judgments would result in estimates close to the base-rate of 1%. Kahneman and Tversky (1972), however, found the opposite result: people seemed to base their judgments only on the new information, a reasoning strategy they coined "representativeness". At the same time, people disregarded the prior probability of the hypothesis, a phenomenon termed the "base-rate fallacy". Judgments by representativeness would result in estimates of 80%, because they fail to integrate the base-rate into the judgment.

The volume of literature published on biases in Bayesian reasoning led Bar-Hillel (1980) to conclude that "the genuineness, the robustness, and the generability of the base-rate fallacy are matters of established fact" (p. 215). Kahneman and Tversky (1972, p. 450) themselves noted that "in his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all", lending little hope for human's capacity to perform correct Bayesian inferences.

More recently, however, studies have emphasized the issue of the representation of information in the reasoning problems: while conditional probabilities as shown above elicit only few correct Bayesian inferences, providing information in natural frequency format significantly increases the number of correct estimates of posterior probabilities (Gigerenzer & Hoffrage, 1995; Hoffrage, Gigerenzer, Krauss, & Martignon, 2002; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000; Sedlmeier & Gigerenzer, 2001). Unlike conditional probabilities, natural frequencies contain information about a set of events or objects, so that base-rate information is already included. To illustrate, using the example from above, the same information in natural frequencies would yield:

1. 10 out of every 1000 women at the age of 40 who participate in routine screening have breast cancer
2. 8 out of every 10 women with breast cancer will get a positive mammography result (H&D)
3. 95 out of every 990 women without breast cancer will also get a positive result in their mammography (¬H&D)

With natural frequencies, the Bayesian calculation reduces to:

$$p(D|H) = \frac{H\&D}{H\&D + \neg H\&D} = \frac{8}{8+95} = \frac{8}{103} \qquad (2)$$

Thus, the number of women who actually have breast cancer out of those who receive a positive test result is 8 out of 103, or 7.8%.

Two factors have been proposed to explain why natural frequencies lead to higher levels of correct inferences compared to conditional probabilities: 1) the computational facilitation natural frequencies provide, and 2) the ecological design of the information, i.e. the resemblance of natural frequencies with information sampling in natural environments (Gigerenzer & Hoffrage, 1995).

**Computational facilitation**

The first explanation proposed for the higher rates of Bayesian solutions with natural frequencies compared to conditional probabilities or relative frequencies is the facilitation of Bayesian computations (Gigerenzer & Hoffrage, 1995; Hoffrage & Gigerenzer, 1998). Natural frequencies reduce the number of necessary computational steps, as a comparison of Equations (1) and (2) reveals, from three multiplications, one addition, and one division with conditional probabilities to one addition with natural frequencies. The reason for this facilitation is that natural frequencies already contain information about the base-rates of events. Revisiting the mammography example above, conditional probabilities are normalized at the point of "breast cancer" or "no breast cancer". They state that 80% *of the women who have breast cancer* will receive a positive test result in the mammography, and that 9.69% *of the women who do not have breast cancer* still test positive. In order to derive correct posterior probabilities, information about how likely a woman has breast cancer has to be considered and integrated. Natural frequencies, on the other hand, are not normalized and represent conjoint events: they indicate the number of women who have breast cancer *and* test positive, which already includes information about the base rate. As a consequence, base rate information does not need to be mathematically integrated, resulting in fewer computational steps. Reducing the number of necessary computational steps also reduces possibilities for error, which is in line with higher rates of correct solutions with natural frequencies.

**Ecological design**

Even with the transparent format of natural frequencies, still not all participants give correct answers. If this is a problem of external quantity representation rather than internal reasoning operations, representing information the way it is encountered in natural environments should increase performance.

In addition to simplifying the computation, Gigerenzer and Hoffrage (1995) argue that natural frequencies have an advantage over conditional probabilities because they mimic the way humans and other animals have encountered probabilistic information throughout the course of their evolutionary history: as single, countable events observed over time, or objects laid out in space. Presenting information in laboratory experiments in a way that conserves the information structure of the environment has been termed ecological design, and goes back to the idea of "representative design" in human perception by Egon Brunswik (1955).

One way of simulating a natural information structure in laboratory experiments is making subjects experience the events sequentially. Humans track the frequency of occurrence of events without explicit instruction, and can judge the frequency of events relatively accurately (Hasher & Zacks, 1984; Zacks & Hasher, 2002). When subjects can sample the events sequentially instead of just receiving a description, Bayesian inferences conform closer to the norm of probability theory (Betsch, Biel, Eddelbüttel, & Mock, 1998; Gigerenzer, Hell, &

Blank, 1988). Bayesian information processing in natural environments is not reserved for humans: ethologists have described foraging behavior in bumblebees and pigeons which indicates complex Bayesian information processing when selecting where to search for food (Real, 1991; Real & Caraco, 1986)

> "The assumption we started out with was that there would be inductive reasoning mechanisms that represent information as frequencies because in our natural environment that is what we would have been encountering: a series of real, discrete, countable events. If true, then the highest levels of Bayesian performance should be elicited when subjects are *required* to represent the information in the problem as numbers of discrete, countable individuals." (Cosmides & Tooby, 1996, p. 33).

Cosmides and Tooby (1996) tested various conditions where the information was presented as icon arrays, which are series of single, countable elements. They found that up to 92% of their subjects could indicate the correct posterior probability with icon arrays, compared to 76% with natural frequencies. It is important to note that performance does not improve with additional graphical information per se: (2009) compared icon arrays and natural frequency formats to other types of graphical information, and found that only those formats improved understanding that contained single countable elements.

Clearly, icon arrays have an advantage over numerical formats, such as natural frequencies, in eliciting correct Bayesian judgments. As will be shown next, icon arrays represent analogue quantities, which can even be comprehended by infants, while numerical formats represent information as visual symbols, whose significance has to be learned in the course of formal education. The reason for the advantage of icon arrays over numerical formats may thus rely in different representations of quantity.

## Quantity Representations in Humans and other Animals

At the beginning of the 20th century, a horse named Hans made it to the headlines of German newspapers. His owner, Wilhelm von Osten, claimed that Hans was able to spell out words, to count, and to solve arithmetic problems, earning him the name "Clever Hans" (cf. Fernald, 1984). Hans demonstrated his ability in van Osten's yard, where the public would gather in a semi-circle and suggest arithmetical questions to van Osten. Hans then solved the questions posed either verbally or written on a blackboard by striking his hooves to give the answer. Remarkably, Hans could solve additions of natural numbers, additions of fractions, and could even indicate the divisors of the number 28. News about the ingenious horse spread fast, and German psychologist Carl Stumpf and his student Oskar Pfungst as members of a scientific committee began to systematically analyze Hans' performance. Their discovery was that not mathematical knowledge allowed Hans to answer the arithmetic problems but rather Hans' precise observation of unintentional signals of his master von Osten. This has sparked research on experimenter expectations and unintentional signaling and learning, but unfortunately discredited claims about numerical competencies in animals.

Animals do, however, possess elaborate mechanisms for processing quantity[1] information. Crows, for example, can differentiate groups of objects based on their quantity (Koehler, 1950; Mechner, 1958), and chimps consistently choose the larger of two piles of chocolates, if the difference in quantity between the two piles is large enough (Rumbaugh, Savage-Rumbaugh, & Hegel, 1987)

To little surprise, humans can also assess quantity from a set of objects and differentiate the quantities of sets of objects. Quantities up to four can be assessed rapidly and mostly error-free, a phenomenon termed subitizing after the Latin *subitus*, meaning sudden (Kaufman & Lord, 1949). Subitizing was first described by W. Stanley Jevons (1871). He observed that he could quickly name the number of up to four black beans in a paper box without making errors, and that errors increased with the number of beans in the box. More recent investigations (e.g., Mandler & Shebo, 1982) found that when asked to name the quantity of objects in a set, response times (RTs) for one, two, or three objects were similar, and increased linearly by 200-300 msec with each additional object after three, while error rates also started to increase for more than three objects. Three or four is therefore regarded to be the maximum number of objects that can be attended to simultaneously. Larger quantities can only be approximated or need to be counted.

Humans are not left with a rudimentary sense for quantity: we use number words and symbols to denote quantities, which enables us to enumerate

[1] By quantity, I will refer to the abstract notion of the number of constituent parts in a set of objects or events, such as five cars in a street or two mentions of the word "cat" in a sentence. In the literature, magnitude (e.g., Dehaene, Bossini, & Giraux, 1993), number (e.g., Xu & Spelke, 2000), and numerosity (e.g., Atkinson, Campbell, & Francis, 1976; Emmerton, 1998) have been used to denote the same concept.

quantities precisely. Dehaene (1997; S Dehaene, Molko, Cohen, & Wilson, 2004) proposed a triple-code-model, where quantities can be encoded either in an analogue form (••), as verbal (*two*), or visual (*2*) symbolic codes. Dehaene's model of development of numerical concepts proposes that children possess a core system for representing analogue quantities in early infancy, which allows them to subitize smaller, approximate lager, or compare groups of objects. In preschool age, children learn to connect analogue quantities to verbal expressions, building a verbal number system required for counting or arithmetic fact retrieval. Finally, Arabic numerals are introduced around school age, allowing for written calculations.

In human development, an analogue representation for quantity is already present in early infancy. In a preferential-looking paradigm, six-month-old children showed longer looking times for groups of dots which differed from an habituated display by the number of dots, controlling for other variables like dot size and luminosity, demonstrating the ability to attend to quantity information in a group of objects (Xu & Spelke, 2000). In another study (Lipton & Spelke, 2003), the discriminability of different stimulus quantities showed a characteristic Weber function, i.e. not the absolute difference between the two sets of stimuli, but the difference relative to the size of the set determines whether a difference can be perceived. While for children at the age of six months the difference between the two sets needs to be quite large (8 vs. 16 stimuli), nine-month-old children can already perceive smaller differences (8 vs. 12 stimuli). The order in which children become familiar with different forms of quantity representation leaves traces in the quantity information is processed by the cognitive system – analogue quantities are processed automatically earlier in development than symbolic quantities. Girelli, Lucangeli and Butterworth (2000) used an adapted version of the Stroop task to examine the interference between physical size and denoted quantity of Arabic numerals in children from first to fifth grade. The children had to judge which of two Arabic numerals was larger in physical size while they also varied in denoted quantity. In adults, the RT to the task depends on the size-quantity combination: size judgments are facilitated if larger quantities are associated with larger symbols, and an interference effect is found if smaller symbols denote larger quantities. For the children studied, this interference effect depended on the age studied. Only older children and adults were influenced by the quantity represented by the Arabic numeral. Third-graders showed an emerging effect of interference and no effect was observed for first-graders. The association of larger quantities with larger symbols seems to develop in the first years of elementary school – although cultural differences are possible. (2007) present evidence that Chinese children already encode quantity from number symbols at the age of five. Gebuis, Cohen Kadosh, de Haan and Henik (2009) examined the association of quantity and physical size with different quantity representations, analogue quantities (groups of dots) and Arabic numerals, for preschool children and adults. As in the study of Girelli et al., adults showed an interference or facilitation effect for Arabic numerals, but children did not. However, if children were asked about the physical size of dots varying in quantity, they also showed an association of physical size with quantity comparable to adults. These results indicate that quantity information is extracted from analogue material without instruction before the same occurs for Arabic number symbols. Also, children can perform basic arithmetic operations like additions and subtractions with sets of objects before they can solve the same problems with Arabic number symbols (Barth, La Mont, Lipton, Kanwisher, & Spelke, 2006).

Analogue representations of quantity also seem to determine mathematical proficiency. Halberda, Mazzocco and Feigenson (2008) report the results of a longitudinal study, which followed children from preschool to ninth grade and found that the ability to discriminate between analogue quantities in ninth grade covaried with scores in standardized tests on mathematical achievement through elementary school and as early as kindergarten age. A more precise analogue representation of quantity was linked to a better understanding of mathematical concepts.

Besides normal interindividual variations in the precision of quantity representations and mathematical aptitude, various factors can lead to specific impairments of numerical competencies in adults, or developmental deficiencies confined to the number domain in children.

First, adults may lose their arithmetic abilities due to lesions in certain regions of the brain. Lesions in the vicinity of the junction of the occipital, temporal, and parietal lobes can cause a multitude of symptoms summarized as Gerstmann syndrome, including finger agnosia, confusing right and left orientation, agraphia, and severe impairments in elementary quantity representations and arithmetic competencies (Gerstmann, 1940). This brain region, referred to as the horizontal segment of the intraparietal sulcus (HIPS), has been located as a critical structure for quantity information. In functional magnetic resonance imaging studies (fMRI), tasks involving number words, number symbols, or analogue quantities all lead to an increased activation in this area of the brain (cf.

S Dehaene, Piazza, Pinel, & Cohen, 2003). Disrupting neural activity by applying transcranial magnetic stimulation (rTMS) selectively impairs processing quantity information, independent of modality (Cappelletti, Barth, Fregni, Spelke, & Pascual-Leone, 2007; Sandrini, Rossini, & Miniussi, 2004).

Second, the developmental model of quantity representations proposed by Dehaene is vulnerable to influences from developmental deficiencies, leading to specific impairments (see von Aster & Shalev, 2007). Some children may not develop adequate representations of quantity to begin with, a condition called developmental dyscalculia.

If the children fail to develop an adequate representation of analogue quantities, number words can still be learned by rote memory, but do not contain any semantic associations to quantity concepts. These children show severe impairments with both, analogue and symbolic quantities (Hanich, Jordan, Kaplan, & Dick, 2001).

On the other hand, if analogue representations are preserved, but linguistic abilities are perturbed, the association of analogue quantities and their verbal representation cannot be established in an age-appropriate manner. Those children exhibit a normally developed analogue quantity representation, but show impairments in handling verbal or visual symbolic quantities. The same may be true for children with attentional or working memory deficits (e.g., ADHD, Geary, Brown, & Samaranayake, 1991). They may have difficulties retrieving and producing counting sequences, or retrieving the correct order of number word elements, such as “one hundred and forty eight”.

## Assessing Bayesian Reasoning Performance

With one exception (Cole & Davidson, 1989), previous research has exclusively focused on rates of correct solutions to Bayesian reasoning problems, and relied on the indicated probability or proportion to infer reasoning strategies. Analyzing the time it takes people to make a judgment or reach a decision is another established method to understand and differentiate psychological processes (e.g., Busemeyer & Townsend, 1993). RT analysis dates back to the work of Franciscus Cornelius Donders (1868, 1969) and Wilhelm Wundt (1880), and became a mainstream tool with the advent of the information processing approach (Sternberg, 1966, 1969). As will be explored later, claims about computational facilitation and ecological design do not only have implications for the rates of correct solutions on Bayesian reasoning tasks with different information formats, but also about the processing times of that information. The study by Cole and Davidson (1989) examined performance and RTs for Bayesian reasoning problems in subjects who were either naïve with regards to Bayesian reasoning problems or had been trained to convert the information from conditional probabilities to one of various graphical formats. They were then tested on problems in conditional probability format. While training with icon arrays led to the best performance and no training to worst, the authors reported fastest RTs in the group without training, and slower RTs for the training conditions. However, RTs were not separated for correct and incorrect responses, which does not allow for a comparison of the speed of Bayesian reasoning processes between the formats, nor a differentiation of reasoning strategies within one format. Untrained subjects could have simply guessed the answer, leading to faster response times but also more wrong answers. Also, as all subjects solved Bayesian reasoning problems with conditional probabilities, their study also does not allow inferences about the possible differences in cognitive processes with different representations of the problem. One focus of this dissertation will be the analysis of RTs as indicators of reasoning processes in Bayesian judgments with different information formats.

Another focus of this work is assessing the level of detail of Bayesian judgments. People are generally better at remembering the imprecise main statements rather than the precise verbatim details, and seem to act primarily on imprecise evaluations (Reyna & Brainerd, 1994, 1995). In a Bayesian reasoning task, judging whether more afflicted persons or more healthy persons will receive a positive test result is a categorical judgment that does not require calculating a posterior probability, but still indicates Bayesian reasoning. For icon arrays and natural frequencies, categorical judgments only require determining which number is higher – H&D or ¬H&D. With conditional probabilities, categorical judgments require comparing $p(H) \times p(D|H)$ to $p(\neg H) \times p(D|\neg H)$, which contains solving two multiplications and is hence more computationally demanding.

## Hypotheses

The main proposition of this thesis is that Bayesian information processing is not intricately difficult for the human mind, but that the ease and speed of Bayesian judgments depends largely on the external representation of information and the differences in the quantity and quality of processing steps that is implied by the external representations. The following chapters will test three predictions that can be derived from this assumption.

First, children at the beginning of formal schooling, with little experience with Arabic numerals, should already show systematic Bayesian thinking when the information is presented in a format that mimics information in natural environments. Also, older children should show a better performance if the information is presented as analogue quantities, compared to numerical material. These hypotheses will be tested in chapter II, including a replication with minimal modifications to the task phrasing.

Second, chapter III will investigate if children with developmental dyscalculia can profit from icon arrays when performing Bayesian reasoning. Based on the developmental model of Dehaene (1997) and von Aster and Shalev (2007), two groups of dyscalculic children exist: one that may grasp analogue quantities, but lacks verbal and visual symbolic representations, and one group where even the most basic, analogue representation of quantity is impaired. While icon arrays should yield higher rates of correct answers to Bayesian reasoning problems compared to numerical information, they should be more beneficial for children who have a preserved sense of analogue quantities, contrasted to children who do not.

Finally, chapter IV is devoted to cognitive processes in adults. The first part of this chapter investigates if differences in performance between Bayesian reasoning problems with different information formats also hold for RTs. The second part investigates if Bayesian reasoning operations rely on different aspects of working memory, according to the mathematical operations implied by the information and response formats.

# II – Bayesian Reasoning in Elementary School

While research about children's ability to engage in deductive reasoning has flourished under the influence of Jean Piaget (e.g., 1971), little is know about how children draw conclusions about the world based on uncertain information and their prior knowledge. Piaget describes cognitive development as the transition from unsystematic and erroneous inferences about the physical nature of the world towards a consistent application of deductive logical principles. The world, however, is not consistent and logical, but stochastic: random variations are the rule, rather than the exception, making information futile and uncertain. Only little is known about how children's thinking is adapted to this fundamental uncertainty.

A previous study by Zhu and Gigerenzer (2006) found that children as early as fourth grade showed correct answers on Bayesian reasoning problems with natural frequencies, and sixth-graders performed on a level comparable to adults with conditional probabilities. The authors also analyzed the children's reasoning strategies and found little support for the use of shortcut-strategies previously described by the heuristics and biases program (cf. Tversky & Kahneman, 1974). Instead, the development from fourth to sixth grade seems to follow an overlapping wave model, with a transition from guessing behavior to focusing on the evidence, to correct Bayesian inferences.

In another study (Girotto & Gonzales, 2008), children as young as four years could use cues and the prior distribution of events to make advantageous choices when betting on uncertain outcomes. Here, they were asked to bet on the outcome of a single event, i.e. the result of one random draw. In contrast to single events, judgments about the frequency of events require judging how often a certain event will occur in multiple random draws. Instead of just asking for a binary judgment such as which event will more likely or more frequently occur, children may be asked more precisely how likely or how often as certain event would instantiate. Unfortunately, the precision in making Bayesian judgments about single events or frequencies of events and the strategies children used were not reported in this study.

The present chapter aims to answer the following questions:

1. Do children profit from information as single, countable elements, such as icon arrays, compared to natural frequencies when solving Bayesian reasoning tasks?
2. How much easier to answer are categorical questions compared to questions about exact posterior probabilities?
3. Do children differentiate between judgments about the frequency of events and about single event probabilities?
4. Can this study replicate the strategies reported by Zhu and Gigerenzer (2006) for fourth-graders, and which strategies do younger children use?

## Methods

### Sample

91 second- and 85 fourth-graders from public elementary schools in Berlin were recruited for this study. The mean age of the second-graders was 7;6 years

(range from 7;1 to 10;0), the mean age of fourth-graders was 9;6 (range 9;0 to 11;1).

**Materials**

We constructed six Bayesian reasoning problem with content suited for children. These problems involved making inferences, based on a prior distribution and a binary cue. One problem is given below, the others can be found in Appendix A.

Sample problem:

> "Out of every 30 men living in a small village, 5 of them work as firemen, and the other 25 do not work as firemen. 4 of the 5 firemen wear a moustache. Also, 15 of the 25 men who do not work as firemen wear a moustache."

This was the only information provided to children in the natural frequency condition. Children in the icon array condition additionally received a graphical display of the information in addition (see Figure 1).

*Figure 1*. Icon array depicting the 30 villagers, some wearing a moustache, and some wearing a helmet, for children in the icon array condition.

After the children looked at the information, they were asked three questions:

1. "If I accidently meet someone from that village who is wearing a moustache, do you think he is a fireman?" with the answering options "rather yes" and "rather no".

2. "Now think of all the men wearing moustaches. Are there more firemen, or more men who do not work as firemen?" where children could answer "more firemen" or "more men who do not work as fireman".

3. “Of all the men with a moustache, how many are firemen? _____ out of _____” where children had to fill in the blanks with their answer.

The first two questions ask for a categorical judgment about a single event (question 1) or the frequency of events (question 2), the third question has to be answered with a numerical estimate of the posterior probability.

**Procedure**

Children were tested in their classrooms in small groups of two to six. Children in each classroom were randomly assigned to either the natural frequency or the icon array condition and solved all six Bayesian reasoning problems. The experimenter read aloud one sample problem, and then gave the correct answers to the questions and a short explanation. All children received the same instruction: “Now you have to solve similar problems on your own. Read each problem carefully and answer all three questions. If you do not know the answer, pick what is most plausible to you.”

A strict coding criterion was used to identify Bayesian responses, i.e. the ratio of the numbers given in question 3 had to match the posterior probability exactly to qualify as a Bayesian response (see Zhu & Gigerenzer, 2006). We used this method because it minimizes the chance of erroneously classifying the result of guessing or the use of a Non-Bayesian strategy as correct inference, although it misses correct inferences which were distorted by calculation errors.

## Results

### Do children profit from icon arrays, compared to natural frequencies only?

In second grade, children answered 11% of all problems correctly with natural frequencies only, and 22% when they had icon arrays in addition ($\chi^2(1, N = 91) = 11.90$, $p < .001$, $\phi = 0.15$; see Figure 2).

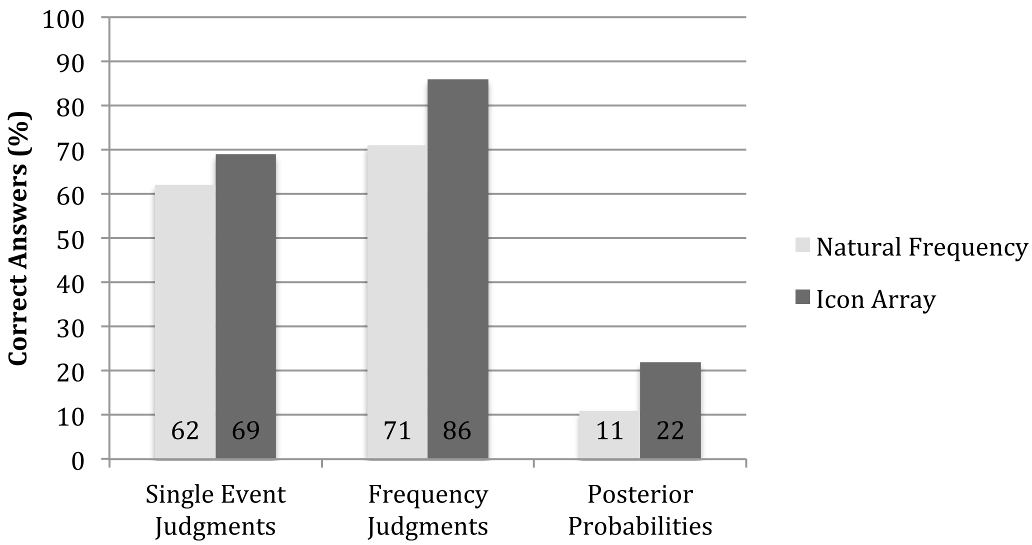



*Figure 2*. Percentage of correct answers in both experimental conditions for judgments about single events, the frequency of events, and exact posterior probabilities for children in second grade.

For fourth-graders, correct answers increased from 42% with natural frequencies only to 60% with added icon arrays ($\chi^2(1, N = 85) = 17.72$, $p < .001$, $\phi = 0.19$; see Figure 3).

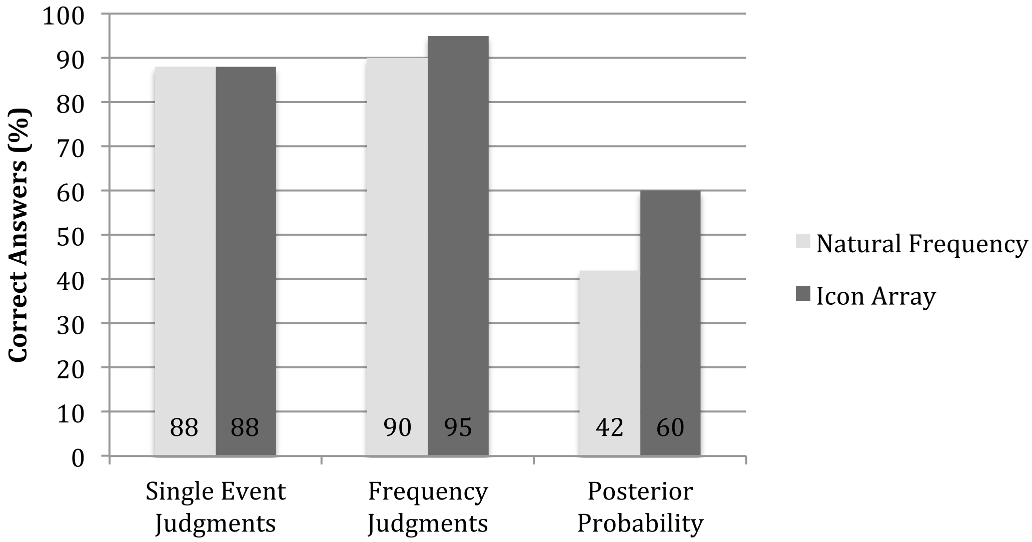



*Figure 3*. Percentage of correct answers in both experimental conditions for judgments about single events, the frequency of events, and exact posterior probabilities for children in fourth grade.

The top section of Table 1 shows that Bayesian answers were not distributed normally, but shifted from left to right from second to fourth grade. In the natural frequency condition second graders exhibit a left-skewed distribution with a mode of 0 correct answers, which falls off steeply. Remarkably, two second-graders were able to answer four or more problems correctly. In the icon array condition, the distribution is bimodal, with peaks at one and four correct answers. In fourth grade, answers in the natural frequency condition were also distributed in a bimodal fashion, with modes of 0 and 4. With icon arrays, most children gave four correct answers.

*Table 1*. Number of responses allocated to one of the strategies for second- and fourth-graders in both experimental conditions. E.g., 8 children in second grade in the natural frequency condition answered 1 problem correctly. Shaded areas indicate systematic strategy selection. The total designates the total number of answers attributed to each reasoning strategy.

| | | Number of Bayesian Responses | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| Second Grade | | | | | | | | | |
| Natural Frequency | 48 | 255 | 8 | 8 | 0 | 1 | 1 | 0 | 33 |
| Icon Array | 43 | 200 | 12 | 6 | 2 | 7 | 0 | 0 | 58 |
| Fourth Grade | | | | | | | | | |
| Natural Frequency | 40 | 139 | 6 | 4 | 5 | 7 | 4 | 4 | 101 |
| Icon Array | 45 | 106 | 2 | 6 | 5 | 17 | 11 | 2 | 164 |
| | | Number of Conservatism Responses | | | | | | | |
| | n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| Second Grade | | | | | | | | | |
| Natural Frequency | 48 | 255 | 14 | 5 | 3 | 0 | 0 | 0 | 33 |
| Icon Array | 43 | 244 | 8 | 3 | 0 | 0 | 0 | 0 | 14 |
| Fourth Grade | | | | | | | | | |
| Natural Frequency | 40 | 229 | 7 | 2 | 0 | 0 | 0 | 0 | 11 |
| Icon Array | 45 | 259 | 7 | 2 | 0 | 0 | 0 | 0 | 11 |
| | | Number of Representative Thinking Responses | | | | | | | |
| | n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| Second Grade | | | | | | | | | |
| Natural Frequency | 48 | 226 | 24 | 10 | 6 | 0 | 0 | 0 | 62 |
| Icon Array | 43 | 217 | 24 | 7 | 1 | 0 | 0 | 0 | 41 |
| Fourth Grade | | | | | | | | | |
| Natural Frequency | 40 | 139 | 11 | 6 | 2 | 3 | 0 | 0 | 101 |
| Icon Array | 45 | 239 | 14 | 7 | 1 | 0 | 0 | 0 | 31 |

*Table 1 (continued)*

| | n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|---|---|
| | | Number of Evidence-Only Responses | | | | | | | |
| Second Grade | | | | | | | | | |
| Natural Frequency | 48 | 265 | 11 | 6 | 0 | 0 | 0 | 0 | 23 |
| Icon Array | 43 | 232 | 16 | 5 | 0 | 0 | 0 | 0 | 26 |
| Fourth Grade | | | | | | | | | |
| Natural Frequency | 40 | 218 | 14 | 4 | 0 | 0 | 0 | 0 | 22 |
| Icon Array | 45 | 256 | 6 | 4 | 0 | 0 | 0 | 0 | 14 |

| | n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|---|---|
| | | Number of Pre-Bayesian Responses | | | | | | | |
| Second Grade | | | | | | | | | |
| Natural Frequency | 48 | 288 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Icon Array | 43 | 256 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Fourth Grade | | | | | | | | | |
| Natural Frequency | 40 | 235 | 3 | 1 | 0 | 0 | 0 | 0 | 5 |
| Icon Array | 45 | 268 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |

| | n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|---|---|
| | | Number of A / C Responses | | | | | | | |
| Second Grade | | | | | | | | | |
| Natural Frequency | 48 | 269 | 11 | 4 | 0 | 0 | 0 | 0 | 19 |
| Icon Array | 43 | 230 | 9 | 3 | 3 | 1 | 0 | 0 | 28 |
| Fourth Grade | | | | | | | | | |
| Natural Frequency | 40 | 221 | 11 | 4 | 0 | 0 | 0 | 0 | 19 |
| Icon Array | 45 | 257 | 11 | 1 | 0 | 0 | 0 | 0 | 13 |

**How much easier to answer are categorical questions compared to questions about exact posterior probabilities, and do children differentiate between judgments about the frequency of events and single-event probabilities?**

Since we used binary forced-choice questions, a guessing strategy would result in 50% correctly answered questions, making this the baseline for evaluating performance. Across grades, experimental conditions, and question formats we observed data unlikely generated by mere guessing (all $\chi^2 > 13.1$, *p*s <.001 under the null hypothesis of guessing).

We consistently observed higher rates of correct answers for categorical questions than for exact judgments, across all grades, conditions, and categorical judgment formats (see Figures 2 and 3).

The proportion of correct solutions to categorical judgments about single events was compared to that for categorical questions about frequencies. For second-graders, more correct answers were observed for questions about frequencies than about single events, both for the children in the natural frequency condition ($\chi^2(1, N = 96) = 7.25$, $p = .007$, $\phi = 0.11$) and those in the icon array condition ($\chi^2(1, N = 86) = 20.69$, $p < .001$, $\phi = 0.20$). For fourth-graders, the proportion of correct answers to single event or frequency questions was not different for children in the natural frequency condition ($\chi^2(1, N = 80) = 0.36$, $p = .550$, $\phi = 0.03$), but higher for frequency judgments compared to single event judgments for the children in the icon array condition ($\chi^2(1, N = 90) = 10.33$, $p = .001$, $\phi = 0.14$).

*The impact of icon arrays on categorical inferences*

To examine the impact of icon arrays on categorical judgments, it was analyzed if more questions could be answered correctly in the icon array condition compared to the natural frequency condition. For the second-graders, children in the icon array condition answered more categorical judgments correctly when asked about the frequency of events ($\chi^2(1, N = 91) = 16.23$, $p < .001$, $\phi = 0.17$), but not single events ($\chi^2(1, N = 91) = 2.59$, $p = .108$, $\phi = 0.07$). The same pattern was observed for fourth-graders: more frequency judgments were answered correctly with icon arrays (90% vs. 95%, $\chi^2(1, N = 85) = 6.12$, $p = .013$, $\phi = 0.11$), but no difference was observed for judgments about single events (both 88% correct answers).

**Which reasoning strategies do children use?**

Do children in this experiment employ the same strategies used to solve Bayesian reasoning problems described by Zhu and Gigerenzer (2006)? How do icon arrays influence the use of reasoning strategies? Based on the numbers children chose to answer question 3, we categorized their answers to the reasoning strategies previously described in the literature. To simplify, the firemen-problem described above is shown as a tree diagram in Figure 4.

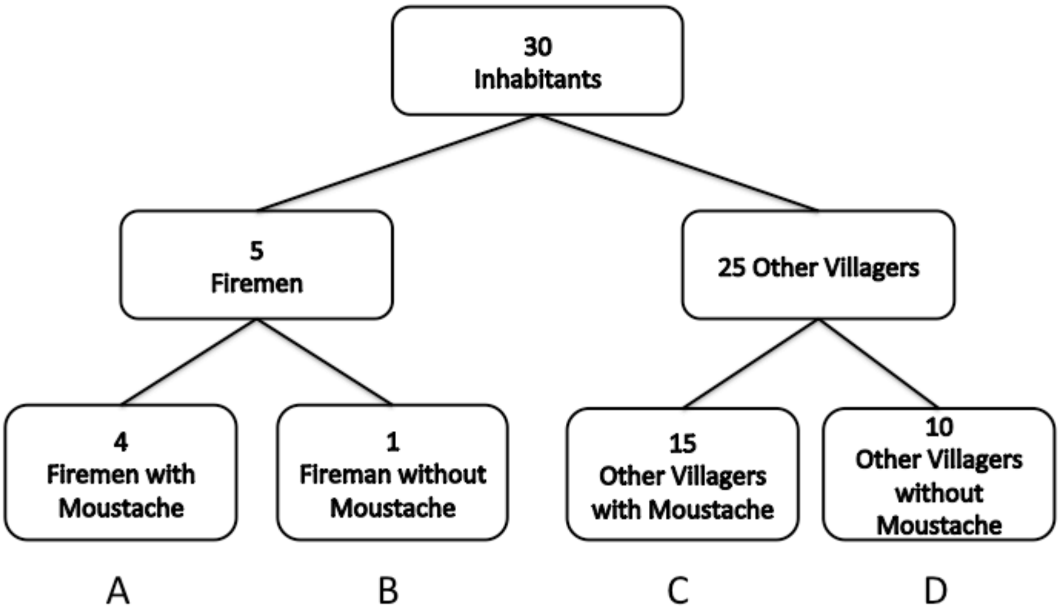



*Figure 4.* The firemen-problem represented as a tree structure. All problems were composed of four elements of information, here labeled A, B, C, and D. Children could combine and use these elements to answer the question about the posterior probability.

*Bayesian reasoning (A/A+C).*

A correct Bayesian inference would be to divide the number of firemen with a moustache (A) by the sum of all people wearing a moustache, regardless of whether they are firemen (A) or do not work as firemen (C). In this example, "4 out of 19" is the correct Bayesian solution.

*Representative thinking (A/A+B).*

Representative thinking describes a misleading reasoning strategy that ignores prior knowledge and only considers the part of the new information that confirms the hypothesis. Judgments are based on the ratio of firemen wearing a moustache (A) to all firemen (A+B), and would produce the answer "4 out of 5".

*Conservatism (A+B/A+B+C+D).*

In contrast to representative thinking, conservatism only takes into account the prior information, ignoring the new information completely. Since there were 5 firemen (A+B) among the 30 inhabitants (A+B+C+D), relying on conservatism would lead to the answer "5 out of 30".

*Evidence-only (A+C/A+B+C+D).*

The intuition underlying this strategy is children focus on new the information and disregard prior knowledge. Following this strategy would result in the proportion of people with moustache (A+C) to all inhabitants in the village (A+B+C+D), leading to the answer "19 of 30".

*Pre-Bayes (A+B/A+C).*

Using a pre-Bayesian strategy means understanding the denominator, but not the numerator of the Bayesian answer. Instead of just counting the firemen with a moustache, children report the ratio of all firemen (A+B) to all people with moustache (A+C), giving the answer "5 out of 19".

As reported, with natural frequencies a Bayesian strategy was used by second- and fourth-graders on 11% and 42% of the problems, respectively. But which strategies were used on the rest of the problems?

*Second-graders with natural frequencies.*

The most commonly used strategies were representative thinking (22% of problems), conservatism (12%), Bayesian reasoning (11%), and the evidence-only strategy (4%). No child in second grade followed a pre-Bayesian strategy, and 51% of the answers could not be classified to one of the previously described strategies. Looking at the responses left unaccounted for, it was analyzed which combinations of numbers were used most frequently and found that the most prominent combination was the ratio A/C – in the example above, the ratio of firemen with moustache to all non-firemen with moustache. This strategy leads to the correct numerator, but lacks the firemen with moustache in the denominator. In total, 7% of the problems were answered in accordance with this strategy. All other combinations of numbers could account for less than 5% of all problems, therefore I will not report them further. Taking into account the A/C-strategy, 45% of all answers could finally not be classified (see Figure 5).

*Fourth-graders with natural frequencies.*

Bayesian reasoning was used on 42% of the problems, while 17% of the problems were answered with the representative thinking strategy, conservatism accounted for 5% of the answers, 7% were answered with the evidence-only strategy, and a pre-Bayesian strategy was used in 2% of the problems. 27% of all answers were left unaccounted for. With 8% of the answers, A/C again was the most commonly used strategy that was not previously described, reducing the percentage of unclassified responses to 19%.

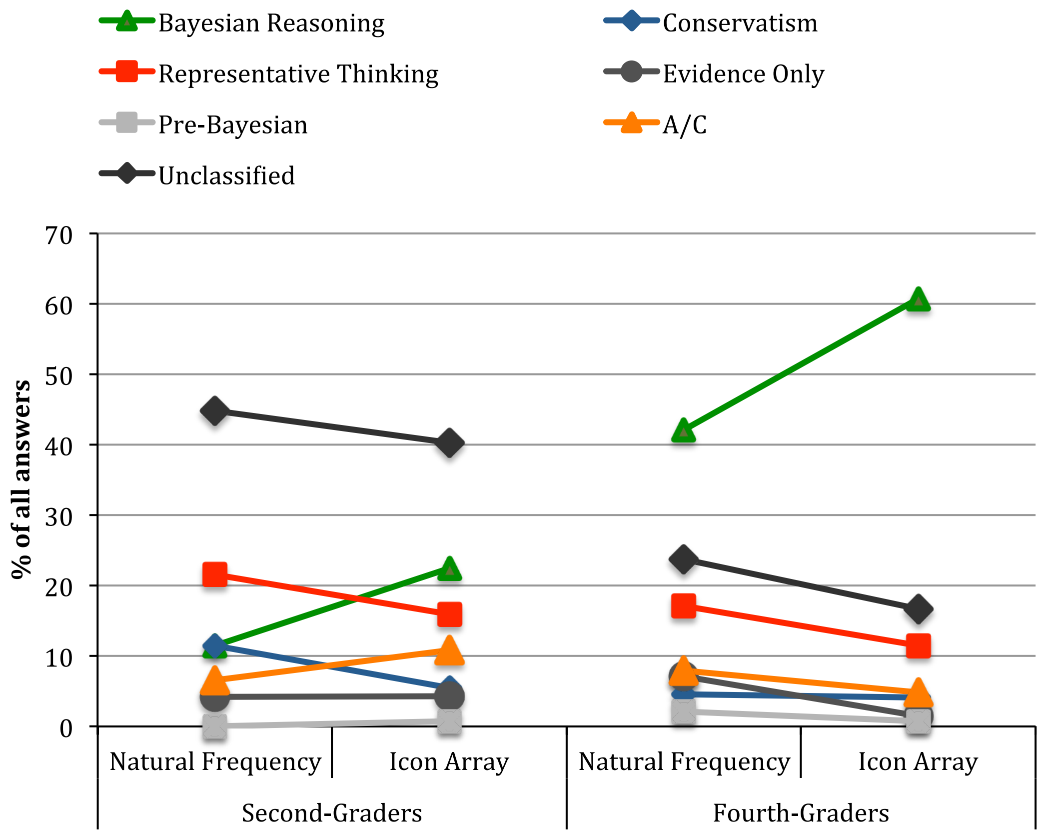



*Figure 5*. The use of Bayesian and non-Bayesian reasoning strategies across grades and conditions. The Figure shows the number of answers allocated to reasoning strategies, based on the numbers children used to answer the problems. Overall, the use of the Bayesian strategy increases from second to fourth grades and when the information is displayed as icon arrays, while the frequency of unclassified answers and the use of non-Bayesian strategies decreases.

**The impact of icon arrays on strategy use**

As reported earlier, icon arrays increase the number of Bayesian judgments, but which strategies are used less frequently as a consequence?

*Second-graders*

Compared to the natural frequency condition, icon arrays increased Bayesian reasoning from 11% to 22% of the problems, and reduced the use of two misleading strategies (representative thinking: 22% vs. 16% of problems, conservatism: 12% vs. 5% of problems). Evidence-only and the pre-Bayesian strategy seemed to be used equally often (evidence-only: 4 vs. 4%, pre-Bayesian: 0% vs. 1%), but the number of problems answered with the A/C strategy increased from 7% to 11%. Finally, the number of unclassified answers also decreased, from 45% to 40%.

*Fourth-graders.*

Children in the icon array condition used all non-Bayesian strategies less frequently than the children in the natural frequency condition (representative thinking: 17% vs. 12%, conservatism: 5% vs. 4%, evidence-only: 7% vs. 2%, pre-Bayes: 2% vs. 1%, A/C: 8% vs. 5%). Unclassified answers were also less frequent with icon arrays than with natural frequencies (17% vs. 19%), while Bayesian answers increased from 42% to 60%. Table 1 summarizes the use of strategies across grades and experimental conditions.

But did the children choose strategies systematically, or were children picking numbers randomly? Following the approach of Zhu and Gigerenzer (2006), we defined three picks of the same strategy as a sign for a systematic use. As the shaded part of Table 1 shows, only two children (4%) in the natural frequency condition in second grade used a Bayesian strategy on three or more of the six problems. In the icon array condition, 21% of the children consistently used a Bayesian strategy, and 16% solved as many as four problems correctly. Interestingly, the consistent use of all non-Bayesian strategies declines with icon arrays, except for the A/C-strategy: no child used this strategy with natural frequencies, but with icon arrays, 9% of the children followed this strategy on three or more problems. Across all conditions and grades, no child used either the evidence-only or the pre-Bayesian strategy consistently.

In fourth-graders, the only systematic use of strategy seems to be the Bayesian strategy and representative thinking With icon arrays, more children followed the Bayesian approach systematically, and only one child (2%) answered three problems using representative thinking.

## Discussion

This chapter tested the hypothesis that Bayesian thinking can be enhanced by using simple, analogue quantity representations such as icon arrays, and that early traces of Bayesian information processing can be detected with answering formats which may be more accessible for children at the beginning of formal education. First, the effect of quantity representations on Bayesian reasoning performance will be discussed, then the debate will focus on exact vs. categorical judgments as answering formats, and finally the results will be related to previous findings on the development of Bayesian reasoning in children.

**Quantity representations**

The literature on the development of quantity representations suggest that an analogue representation enables us to quickly and accurately assess the quantity of a smaller set of objects or estimate the quantity of a larger set of objects, and compare two sets of objects based on their quantity (Stanislas Dehaene, 1997). In the first years of formal schooling, children learn symbolic codes for quantities, and stable associations between Arabic numerals and the magnitude they represent emerge towards the end of elementary school, around the age of 10 years (Girelli et al., 2000). The results show that children can solve more Bayesian reasoning problems if the information is displayed in form of icon arrays. This was observed in the number of problems solved correctly, and the number of children who consistently used a Bayesian strategy, both in second and fourth grade. In second-graders, icon arrays not only produce an increase in the number of correct solutions, but lead to a qualitative change in strategy use – numerical information alone only sparks occasional flashes of Bayesian reasoning with only two children (out of 48) consistently using a Bayesian strategy. With icon arrays, nine children (out of 43) now follow Bayesian logic on three or more problems. This result is consistent with the assumption that not Bayesian reasoning *per se* is the problem, but rather the use of quantity representations which children yet have to automatize. For fourth graders, similar results were observed. However, icon arrays did not lead to the same qualitative change in reasoning strategies oberserved in younger children. Rather, they increased the use of the correct strategy which was already part of the mental toolbox for processing uncertain information.

**Response formats**

Higher rates of correct responses were predicted if the response format does not require arithmetic operations or a numerical answer.

Children in second and fourth grades showed rates of correct solutions for categorical judgments far above those for exact proportions, with fourth-graders performing close to 100%. Children seem to have a sense for the direction of a cue (moustache) given certain prior information (the number of firemen in the village) when asked to make categorical inferences (fireman vs. not a fireman, given someone is wearing a moustache), before they can deliberately apply the correct strategy to determine the exact posterior probability. This finding is consistent with the results of Girotto and Gonzales' (2008) study. Here, children from preschool to elementary school ages had to pick the puppet they thought was most likely to win a piece of chocolate, based on the prior distribution of tokens and new cues. The experimenter then drew a token, but before revealing it to the children, announced the color, changing the odds in favor of the less frequent shape. Children beyond preschool age were able to use this new information and chose the puppet representing the less frequent shape. However, their study did not ask for exact judgments of posterior probabilities, so that the size of the enhancing effect of response format could not be estimated. Our results suggest that the effect can be as large as 64 percentage points – and depends on the representation of quantity, whether the question is about the likelihood of a single event or the frequency of events, and the age of the children.

Girotto and Gonzales (2008) concluded that, based on the above-chance performance of their children for judgments about single events, children's minds were not specially tuned to frequencies, but our results suggest the opposite. While Girotto and Gonzales did not compare judgments about frequencies and single events, the data presented here reveal that although children were generally able to use the information for judgments about single events, especially younger children gave more accurate judgments for frequencies, and even more so when provided with information in an accessible format.

**Reasoning strategies**

While Zhu and Gigerenzer (2006) reported the development of reasoning strategies for children from fourth to sixth grade, to the best of my knowledge, this study is the first to examine Bayesian reasoning strategies with natural frequencies in younger children. In second-graders, flashes of Bayesian thinking were found with natural frequencies, but only two children seemed to follow a Bayesian strategy consistently. Children in fourth grade were able to solve a substantial amount of problems correctly and used a Bayesian strategy systematically when the information was presented in natural frequencies. This result is in accordance with the study of Zhu and Gigerenzer (2006), but a higher rate of correct solutions was observed in the present study. While fourth-graders in the study of Zhu and Gigerenzer solved 17% and 19% of the problems correctly, fourth-graders in this study correctly answered 42% of the problems, a level comparable to Zhu and Gigerenzer's fifth-graders. Two possible explanations may account for this difference. First, the numerical range in our study materials was between 10 and 30, whereas Zhu and Gigerenzer used problems with a reference class of 100. This difference may have made our problems computationally simpler, and hence may have led to more correct Bayesian answers. Secondly, testing in this study was conducted in the second term of the school year, and our fourth-graders were on average three months older than those studied by Zhu and Gigerenzer. Possibly, children in the present study have had more exposure to mathematics classes, enabling them to solve these questions better.

Concerning the use of non-Bayesian strategies, we can confirm results from Zhu and Gigerenzer that in fourth grade, conservatism, evidence-only, and pre-Bayesian strategies are not consistently used and can account for only a small fraction of the answers in the natural frequency and icon arrays condition. However, the number of unaccounted answers in the present study was lower. It appears that children relied less on guessing, and answered problems more systematically. Also, we found that 17% of all problems were answered by following the representative thinking strategy, and 13% of fourth-graders followed this strategy consistently, which is in contrast to Zhu and Gigerenzer's study, which did not find any support for representative thinking. With icon arrays, children answered fewer problems with non-Bayesian strategies, and (except for one child) used only a Bayesian reasoning strategy and A/C systematically. We also observed an increase of A/C-answers. While this strategy does not reflect proper Bayesian reasoning, it is not necessarily misleading but may reflect a comparison of the number of cases in which a cue points in one direction (firemen with moustache) to the number of cases a cue points in the opposite direction (other villagers with moustache). Children who follow this strategy will make correct categorical judgments, but will report an incorrect exact proportion. Since systematic use of this strategy was only observed in second-graders with icon arrays and not the other groups, this strategy should be treated with caution and be subjected to replication before being considered a genuine predecessor of proper Bayesian reasoning in children.

To summarize, in accordance with Zhu and Gigerenzer, children in fourth grade could solve Bayesian reasoning problems with natural frequencies. In contrast to their results, the findings presented in this chapter suggest that fourth-graders do apply misleading non-Bayesian strategies consistently. Second-graders used reasoning strategies less systematically. Presenting the information as icon arrays yielded large effects on the number of correct Bayesian responses. With icon arrays, second-graders showed consistent use of Bayesian reasoning strategies, and children from both grades used misleading strategies less often.

**Conclusions**

Why have previous studies in psychology attested *Homo sapiens* a lack of Bayesian reasoning, while behavioral biologists granted *Columbia livia* – the pigeon – and other animals Bayesian information processing capabilities (Real, 1991)? Goldstein and Gigerenzer (2002) have answered that question with the need to consider both: the mental mechanisms at work, and the environment in which they evolved. This concept has been termed *ecological rationality* and can be described with Herbert Simon's metaphor of the mind as a pair of scissors. One blade represents the cognitive abilities, the other blade represents the environment. Just like we cannot understand the function of the pair of scissors by just studying one single blade, we cannot understand human cognition if we do not consider the structure of the environment in which our cognitive systems evolved. Gigerenzer and Hoffrage (1995) showed that Bayesian reasoning performance can be greatly improved if the information is presented in natural frequencies, because they mimic the way we encounter information in our daily lives. I have taken this point even further and showed that children with little experience with numerical notations can use uncertain information to systematically engage in Bayesian thinking, if the information is provided the way we actually perceive it in our environments – as single, countable elements.

However, this study can be criticized because not all problems involved inferences from the sample described in the information section to another sample, possibly compromising the validity of the results. A replication with the necessary modifications to the question phrasing will be reported in the next part of this chapter.

## Retest

While the goal of chapter II was to assess Bayesian reasoning performance in children at the beginning of formal education, critics may argue that upon close inspection, three of the six reasoning problems used in that study were not about *inferences* from one group to another (problems 1, 5, and 6, see Appendix A). These problems demand a re-description of the information presented (see Table 2).

While the question about single events demanded an inference from the group to an individual, the group of people in the description and the question was identical for frequency judgments. Children did not have to use the information to make an inference about a new group of people. Similarly, the question about the posterior probability did not necessarily demand an inference, but may simply have asked for a re-description of the group, but not necessarily an inference. Possibly, the observed difference between judgments about single events and frequencies of events reported in the first part of this chapter were driven by the different demands both questions pose. Also, the rate of correct Bayesian posterior probabilities may have been inflated because children may have treated the three flawed problems as arithmetic exercises instead of true reasoning tasks. To be precise, if the flawed phrasing of the questions rather than the nature of the inference caused the performance differences observed before, rates of correct solutions for inference questions about single events and frequencies should be the same for problems with correct phrasing. Also, lower rates of correct posterior probabilities should be observed.

*Table 2*. An example of possibly flawed question phrasing used in the first part of chapter II, and the corrected phrasing used in the retest. Bold parts indicate the sample the judgment is aimed at.

| | | |
|---|---|---|
| Problem Text | Out of every 30 people living in a small village, 5 work as firemen, and the other 25 do not work as firemen. 4 of the 5 firemen wear a moustache. Also, 15 of the 25 men who do not work as firemen wear a moustache. | |
| | Flawed phrasing | Corrected phrasing |
| Frequency Judgments | And now think about **all the men wearing a moustache**. Are there more firemen or men who do not work as firemen?” | Now imagine I meet **a group of people from the village who all wear a moustache**. Do you think there are more firemen or more other villagers **in that group**? |
| Posterior Probability | Of **all the men** with a moustache, how many are firemen? | **In that group of people** with moustaches, how many are firemen? |

To test whether the effects observed before were caused by the flawed formulations, the three problematic problems were reformulated to demand inferences (see Appendix B). The corrected phrasing ensured that the sample was not identical to the group described in the problem text, so that children had to make inferences from one group to another. Since for single event judgments, the sample is always different from the group described, the phrasing remained the same. All six problems were tested on a new group of children.

## Methods

### *Sample*

66 children (35 boys and 31 girls) from second and fourth grades of a public elementary school in Berlin participated in the study. The mean age of the children in second grade was 7;6 years (range 6;10 to 9;1), fourth graders mean age was 9;8 years (range 8;1 to 10;10).

### *Materials*

The same problems were used as previously described. However, the wording of three problems was altered to involve an inference to another group than the one described in the introduction text for judgments about the frequency of events and the posterior probability.

### *Procedure*

The procedure was the same as described in the first part of this chapter. Children were tested in small groups of two to six in their classrooms. In each classroom, the children were randomly assigned to either the natural frequency or the icon array condition and solved all six Bayesian reasoning problems. First, the experimenter read aloud one sample problem, and then explained the correct answers to the questions. All children received the same instruction: "Now you have to solve similar problems on your own. Read each problem carefully and answer all three questions. If you do not know the answer, pick what is most plausible to you."

## Results

### *Judgments about single events vs. frequencies*

As the comparison between the number of correct judgments for questions about single events and frequencies reveals, children descriptively gave more correct responses when asked about the frequency of events, although the effects were small. Specifically, children in second grade with natural frequencies only solved 56% of the questions about single events correctly, but 67% of the questions about frequencies ($\chi^2(1, N = 32) = 2.34$, $p = .126$, $\phi = 0.11$). With icon arrays, second graders gave correct single event judgments for 90% of the problems, but could answer 97% of the questions about the frequency of events correctly ($\chi^2(1, N = 30) = 4.87$, $p = .027$, $\phi = 0.16$). Figure 6 summarizes the results of the children in second grade.

Fourth-graders in the natural frequency condition answered 81% of the questions about single event probabilities correctly, compared to 86% for frequency judgments ($\chi^2(1, N = 32) = 1.99$, $p = .158$, $\phi = 0.10$). In the icon array condition, answers for both categorical judgment formats were at the possible maximum (see Figure 7).

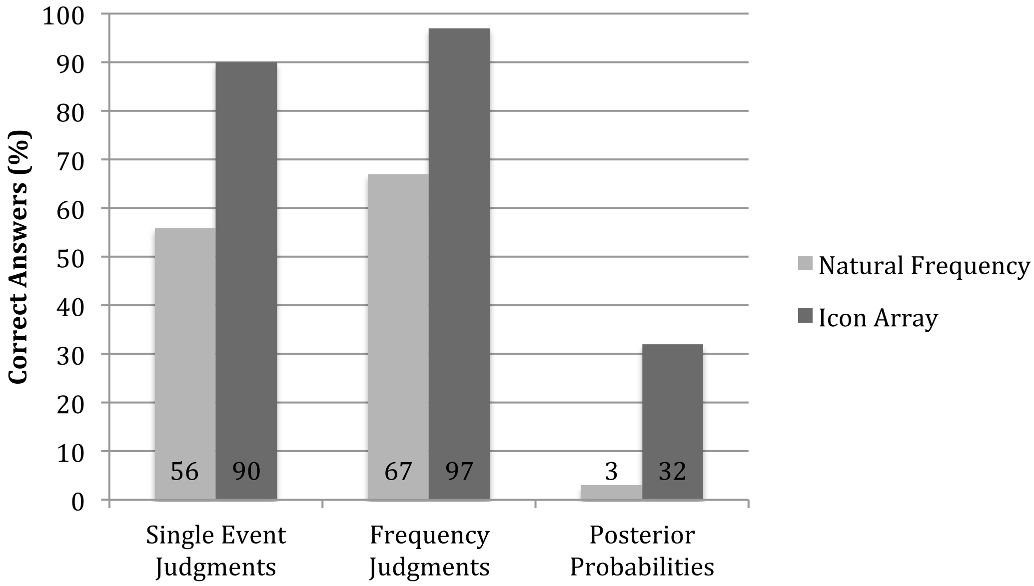



*Figure 6*. The proportion of correct answers for judgments about single events and frequencies, and correct Bayesian posterior probabilities for children in second grade.

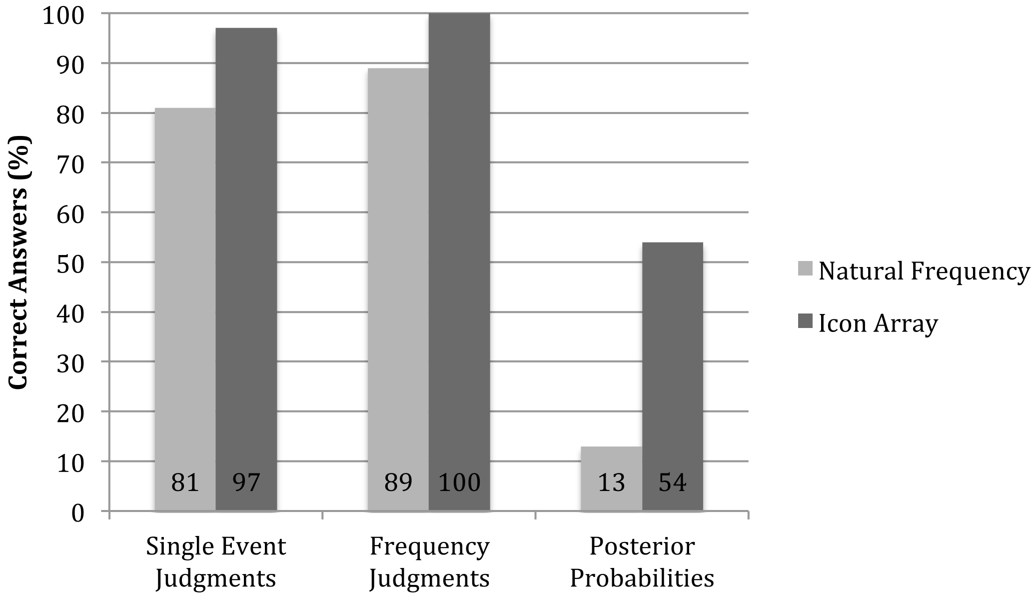



*Figure 7*. The proportion of correct answers for judgments about single events and frequencies, and correct Bayesian posterior probabilities for fourth-graders in both experimental conditions.

*Posterior probabilities*

As Figure 6 shows, second graders with natural frequency information could only indicate the correct Bayesian posterior probability for 3% of the problems, resulting from 3 of the 15 children who answered one problem correctly (see Table 3), compared to 32% correct answers for those children who received icon arrays in addition. A similar picture emerged for the fourth graders: 13% of the problems were solved correctly with natural frequencies alone, and 54% with icon arrays.

*Table 3*. The distribution of correct Bayesian posterior probabilities for both experimental conditions. E.g., three second graders with natural frequencies and icon arrays solved four problems correctly. The total designates the total number of answers attributed to Bayesian reasoning.

| | | Number of Bayesian Responses | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| Second Grade | | | | | | | | | |
| Natural Frequency | 15 | 87 | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| Icon Array | 19 | 90 | 6 | 6 | 2 | 3 | 0 | 0 | 24 |
| Fourth Grade | | | | | | | | | |
| Natural Frequency | 16 | 84 | 4 | 2 | 0 | 1 | 0 | 0 | 12 |
| Icon Array | 16 | 42 | 1 | 1 | 7 | 3 | 2 | 1 | 52 |

**Discussion**

Overall, revising the three items to actually demand inferences still led to higher rates of correct solutions for questions about frequencies, compared to single event judgments, However, the effects were small, suggesting that even second graders can make judgments about frequencies or single events about equally well. Compared to the results presented in the first part of this chapter, single event judgments of second graders showed a marked improvement when icon arrays were presented, from 56% to 90%. Second graders in this study, unlike in the previous, seem to be able to use the information conveyed through icon arrays to make reasonable judgments about single events, while their performance given natural frequencies only was little different from guessing performance. Note that the phrasing of questions about single events was not altered in the retest, so that flawed phrasing cannot account for this result. In comparison to the results in the first part of this chapter, differences between the natural frequency and the icon array condition were even more pronounced. For second-graders, the rate of correct posterior probabilities increased from 3% to 32%. Possibly, icon arrays induce a qualitative shift in reasoning strategies from unsystematic, random guessing about single events and Bayesian posterior probabilities, to a true understanding and application of Bayesian reasoning. This result is compatible to findings from developmental sciences that a stable association between the quantity represented by a numerical symbol and the symbol itself emerges in the first years of elementary school (Gebuis et al., 2009). Along with quantity representations, reading and verbal comprehension skills develop at beginning of formal schooling (e.g., Maliphant, Supramaniam, & Saraga, 1974), which may contribute to the advantage of icon arrays over natural frequency information. While natural frequencies require sufficient reading abilities to comprehend the problem description and the information, icon arrays also alleviate this cultural burden. Future studies on Bayesian reasoning ability in children could investigate the role of icon arrays in overcoming low reading abilities.

To summarize, the results reported earlier were likely not caused by a flawed formulation of the questions. It seems that indeed not Bayesian reasoning operations per se are inaccessible for children, but that children have problems understanding the – verbal or numerical – codes the information is provided in.

# III – Bayesian Reasoning with Impaired Quantity Representations

As chapter II addressed, children who are at the beginning of formal education and are just developing stable connections between different representations of quantity can use icon arrays to systematically draw Bayesian inferences. This chapter is dedicated to Bayesian reasoning abilities in children who – due to genetic and/or developmental deficiencies – did not develop adequate analogue or symbolic representations of quantity, and as a result lack basic arithmetic competencies, summarized as dyscalculia.

A number of possible causes for dyscalculia have been proposed. First, certain genetic abnormalities have been associated with decreased arithmetic abilities, e.g., Tuner's syndrome (Bruandet, Molko, Cohen, & Dehaene, 2004; Temple & Marriott, 1998), Fragile X syndrome (Hagerman et al., 1992), Velocardiofacial syndrome (De Smedt et al., 2007), phenylketonouria (Pennington, 1995), or Williams syndrome (Boddaert et al., 2006). Also, genetic risk factors for dyscalculia have been identified (Alarcon, DeFries, Light, & Pennington, 1997; Shalev et al., 2001). Intrauterine adverse events like alcohol exposure (Kopera-Frye, Dehaene, & Streissguth, 1996), or risk factors such as having a low birth weight (Isaacs, Edmonds, Lucas, & Gadian, 2001) may also contribute to the development of dyscalculia.

Also, developmental delays may occur in the transition from analogue to symbolic quantity representations. This condition, termed developmental dyscalculia, is a specific learning disability affecting the normal acquisition of arithmetic skills. While the etiology of developmental dyscalculia is still being debated, several factors are discussed which may contribute to a poor understanding of numbers. Von Aster and Shalev (2007) propose that developmental dyscalculia may be divided into two subtypes based on the etiology. The first, called *pure developmental dyscalculia* (pDD), results from genetic or intrauterine adverse events which affect the basic understanding of quantity. Children who developed pDD are severely impaired in extracting the quantity from a set of objects and developing symbolic quantity notations like number words or symbols. As a result, they experience problems with both, analogue and symbolic quantities. The second subtype, *comorbid dyscalculia and dyslexia* (cDD), consists of symbolic number processing deficits which may be the consequence of comorbid dyslexia. When regular children learn to associate number words with analogue quantities, children suffering from cDD cannot establish a stable connection between the two concepts. Ultimately, when visual number symbols are introduced, those children can likewise not connect the numeral to a number word. As a consequence, those children possess an intact analogue representation of quantity, while their representations of symbolic quantities are impaired. If the numerical format is the stumbling block on the way to proper Bayesian reasoning in children, rather than the mental mechanisms for quantitative reasoning, icon arrays should improve reasoning performance in children with dyscalculia similar to the effect found in children without learning disabilities. Moreover, the beneficial effect of icon arrays should be larger for children with intact representations of analogue quantities compared to children who suffer from dyscalculia because they lack the basic representation of quantity.

## Methods

### Sample

39 children from second to sixth grades with an existing diagnosis of dyscalculia according to DSM-IV, i.e., unusually low mathematical abilities considering the child's age, intelligence, and education, participated in this study. The children were recruited through their schools or educational therapists. Written informed consent was obtained from the parents prior to collecting the data in accordance with the regulations of the ethics committee of the Max Planck Institute for Human Development. Children received 10€ for their participation. Table 4 summarizes the characteristics of the sample.

*Table 4*. Demographic and psychometric characteristics of the two experimental groups.

| | Icon Array (n = 20) | | Natural Frequency (n = 19) | |
|---|---|---|---|---|
| | M[1] | (SD[2]) | M[1] | (SD[2]) |
| Age | 9;9 | (1;8) | 9;7 | (1;7) |
| IQ | 99 | (13) | 95 | (10) |
| Gender | female: 17 | | female: 14 | |
| | male: 3 | | male: 5 | |
| Comorbidities | ADHD[3]: 1 | | dyslexia: 1 | |

[1] arithmetic mean

[2] standard deviation

[3] attention deficit hyperactivity disorder

### Design and procedure

Each child was randomly allocated to one of two experimental groups, natural frequency or iron array. Children were tested individually in a quiet environment for approximately 45 minutes. First, they solved six Bayesian reasoning problems, followed by the short version of the *Culture Fair Intelligence Test* (CFT 20-R), and finally one subtest of the *Rechenfertigkeiten- und Zahlenverabreitungs-Diagnostikum für die 2. bis 6. Klasse* (RZD 2-6).

*Bayesian reasoning problems*

The reasoning problems used here were modified after those presented in the retest in chapter II. Information was either displayed in numerical format as natural frequencies, or graphically via icon arrays. Here is a sample problem from the natural frequency condition:

> "Out of every 15 pieces of chalk in a box, there are:
> - 4 which are broken and red
> - 1 which is broken and white
> - 6 which are not broken and red
> - 4 which are not broken and white"

Children then saw an icon array of 15 chalk sticks where color and wholeness were ambiguous (Figure 8).

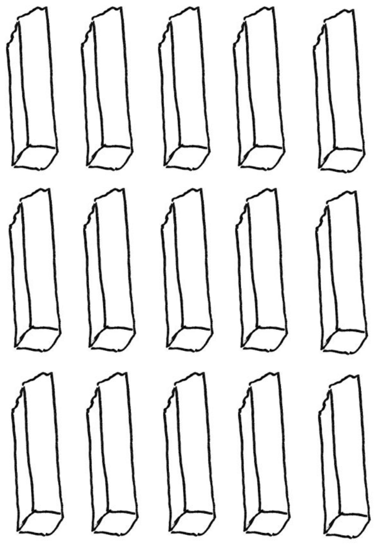


*Figure 8*. Icon array of 15 chalk sticks which does not contain information about the color or intactness of the sticks.

The same problem in the icon array condition would read:

> "A box contains pieces of chalk. Some of the chalk pieces are broken, the other are whole. Some of the broken pieces of chalk are red, and some of the whole pieces of chalk are also red. This is how a box could look like."

This introduction was followed by a picture of pieces of chalk, following the same distribution as in the natural frequency condition (see Figure 9).

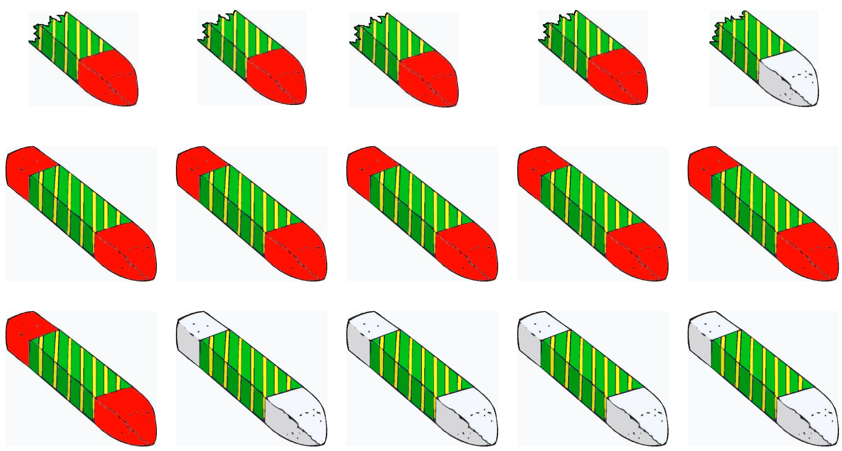


*Figure 9*. Icon array with all necessary information about the color and intactness of the chalk sticks included.

In both experimental conditions, the experimenter would read the introduction text to the children and then asked:

> "If you take some red chalk sticks out of the box, do you think that more are broken, or that more will be whole?"

This question assesses the children's ability to combine the information presented before for categorical judgments. After children gave their answer, they were instructed to mark the chalk sticks:

> "Now look at the picture above. Take a green pencil and circle every red chalk stick in the box. Then take a red pencil and cross out every red chalk stick that would be broken."

Marking the icons in this way was used by Cosmides and Tooby (1996) to measure non-numerical, but more fine-grained Bayesian posterior probabilities. Finally, after the children were done, they were asked:

> "Now, how many of the red chalk sticks that you took out of the box before would be broken? ___ out of ___"

Here, children had to write down their answer, which corresponds to a numerical estimate of the posterior probability. Numerical responses may be problematic in children with dyscalculia, in that using only numerical estimates of posterior probabilities could miss correct Bayesian inferences which were distorted by problems in generating the appropriate numerical output. Bayesian reasoning abilities as assessed by numerical response formats may therefore be underestimated. For that reason, the second question was included, which does not require any numerical knowledge, but yields a graphical description of posterior probability estimates. Like in chapter II, a strict coding criterion was used to identify Bayesian responses. The proportion indicated by the children had to match the posterior probability exactly to minimize the chance of classifying the result of a non-Bayesian reasoning strategy or pure guessing as correct Bayesian inference.

*CFT 20-R*

The CFT 20-R (Weiss, 2006) is a culture-fair and language-free intelligence test, which assesses formal-logical operations and involves recognizing and transforming spatial configurations of stimuli. It was designed for target populations with language comprehension difficulties or outside the western civilization and loads on fluid factors of intelligence. In this study, part 1 was administered, which contains four subtests (completing rows, classifications, matrices, and topological reasoning) and has a retest-reliability of 0.8. Children have a fixed amount of time to solve each subtest, and the number of correct answers is transformed into IQ-scores using age-related norms.

*RZD 2-6*

The subtest „quantity estimation" of the RZD 2-6 (Jacobs & Petermann, 2005) was used to assess the acuity of analogue quantity representations. Children were shown pictures of different objects, varying in quantity, for two (small quantities) or five (larger quantities) seconds. The presentation time is set so that counting the objects is not possible. Children were then asked to indicate the quantity of the objects. Their responses were evaluated based on norms for different school grades, and converted into percentiles.

## Results

Before testing the hypotheses, both experimental groups were compared regarding their demographic characteristics to rule out possible covariate effects. The two experimental groups did not differ in age ($t(35) = 0.48$, $p = .632$), IQ ($t(35) = 0.95$, $p = .350$), performance on the RZD 2-6 "quantity estimation" subtest ($t(35) = 0.93$, $p = .357$), or in their gender distribution ($\chi^2(1, N = 39) = 0.77$, $p = .382$).

#### Icon arrays vs. natural frequencies

To analyze whether icon arrays lead to a better understanding and hence to more correct inferences, icon arrays were compared to natural frequencies for the three dependent variables (see Figure 10).

For categorical judgments, children in the natural frequency group answered 44% of the problems correctly, while the proportion of correct responses in the icon array condition was 78% ($\chi^2(1, N = 39) = 29.35$, $p < .001$, $\phi = 0.35$).

More correct drawings were also found with icon arrays, compared to natural frequencies: 77% vs. 40% ($\chi^2(1, N = 39) = 31.87$, $p < .001$, $\phi = 0.37$).

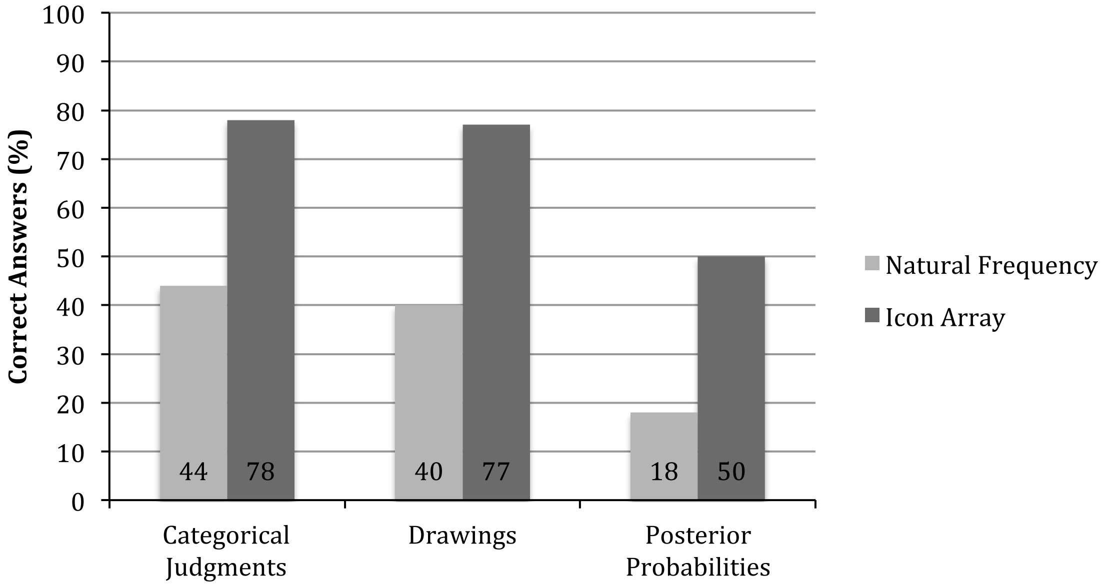



*Figure 10*. Proportions of correct answers, separated for the experimental groups, for the three dependent variables: 1) categorical judgments about the frequency of events, 2) drawing the correct ratio according to the posterior probability in the icon array, and 3) numerical estimates of posterior probabilities.

Finally, children in the icon array condition stated correct Bayesian posterior probabilities for 50% of the problems, while the rate of correct solutions for the children in the natural frequency condition was 18% ($\chi^2(1, N = 39) = 27.37$, $p < .001$, $\phi = 0.34$).

**The role of analogue quantity representations**

Children's performance on the "quantity estimation" subtest of the RZD 2-6 was used to assess anomalies in analogue quantity representation. Children with a percentile of less than 16 were classified as having an impaired quantity representation, as their performance was two standard deviations below their respective grade average. In the icon array condition, eight children were classified as having impaired representation of analogue quantities, the remaining twelve were classified as unimpaired. In the natural frequency condition, nine children were classified as impaired, while the other ten had a quantity representation within the normal range. Comparing the proportions of correct Bayesian posterior probabilities for children with normal and impaired quantity representations yields higher rates for children with normally developed quantity representations in the icon array condition (40% vs. 57%, $\chi^2(1, N = 20) = 3.47$, $p = .062$, $\phi = 0.17$), but no difference for the natural frequency condition (19% vs. 17%, $\chi^2(1, N = 19) = 0.07$, $p = .795$, $\phi = 0.02$; see Figure 11).

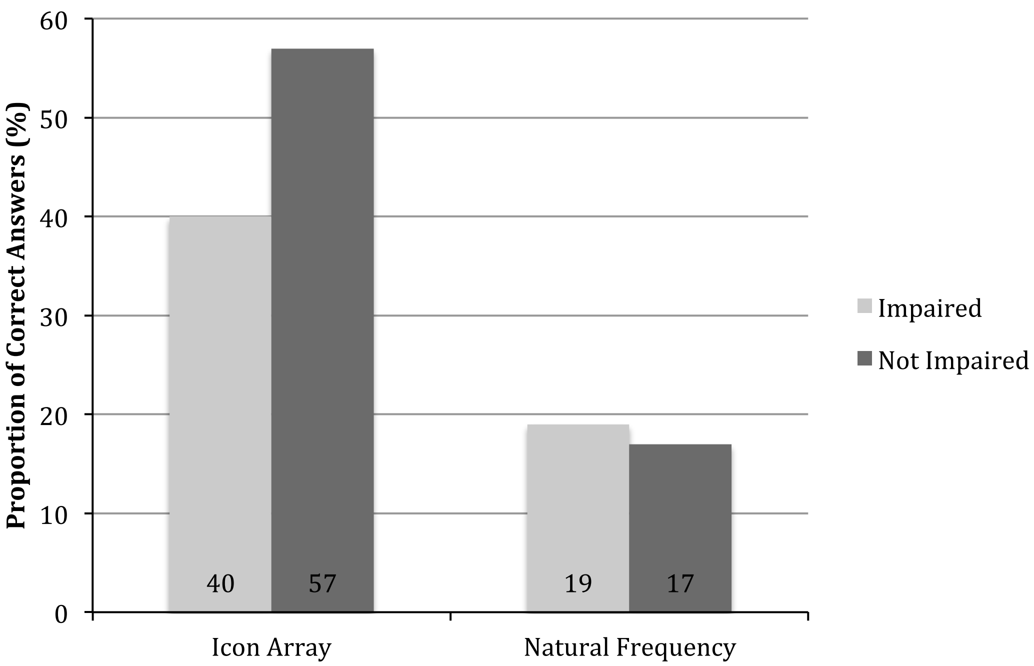



*Figure 11*. Proportion of correct numerical estimates of posterior probabilities for both experimental conditions. Children with impaired representations of analogue quantities showed fewer correct solutions with icon arrays, but not natural frequencies, compared to children with intact quantity representations.

## Discussion

Children with dyscalculia showed a marked increase in Bayesian reasoning performance with icon arrays compared to numerical information, i.e., natural frequencies. This effect was observed for categorical judgments about the frequency of events, as well as for numerical and non-numerical assessments of posterior probabilities. As expected from children with difficulties in number processing, higher rates of correct solutions were obtained if children could draw the proportion representing the posterior probability rather than giving a numerical estimate. However, even when a numerical response was prompted, icon arrays led to more correct responses than natural frequencies.

As expected from the heterogeneous etiologies of dyscalculia (von Aster & Shalev, 2007), icon arrays were less beneficial for children with imprecise mappings of analogue quantities to internal quantity representations. Possibly, those children lack what Dehaene (1997, 2001) called “the number sense” – the ability to perceive and manipulate analogue magnitudes nonverbally on an internal mental number line. If analogue representations of quantities cannot be fully acquired due to developmental or genetic deficiencies, icon arrays cannot provide a large advantage over symbolic quantity representations. However, if children develop dyscalculia because comorbidities such as dyslexia prevent them from learning the connection between number words or numerical symbols and analogue quantities, icon arrays can still be accessed through the intact analogue representation while numerical information cannot be encoded properly.

In conclusion, this study showed the expected advantage of icon arrays over natural frequencies in eliciting correct categorical and exact judgments of posterior probabilities. Children with intact analogue quantity representations benefit more from icon arrays than those with imprecise representations, which is in line with developmental models assuming differential deficits according to the etiology of dyscalculia.

# IV – Response Times and Dual-Task Performance in Bayesian Reasoning

Changing the information from conditional probabilities to natural frequencies does not only bring the information structure of the problem closer to evolutionary reality, it also simplifies the Bayesian calculation. As the comparison between Equations (1) and (2) shows (see. pp. 2-4), the difference in computing exact posterior probabilities between conditional probabilities and natural frequencies consist of three multiplications. As each multiplication alone increases the chances for a mistake and processing time, we do not only expect fewer correct Bayesian answers with conditional probabilities, but also that responses are slower, compared to natural frequencies. Calculating posterior probabilities with natural frequencies requires the addition of two numbers. Response times (RTs) for additions increase with the total sum of the problem, known as the problem-size effect (Ashcraft, 1992). Giving exact posterior probabilities with icon arrays follows the same formula as for natural frequencies, but requires subjects to enumerate the icons via counting to reach a solution. RTs for counting also increase linearly with each item in the set (Atkinson et al., 1976; Mandler & Shebo, 1982), so that we expect slower RTs for icon arrays compared to natural frequencies, and that RTs in both formats co-vary with the total number of icons that need to be counted or the size of the addends, respectively.

Making categorical inferences with conditional probabilities requires to compute almost the full Bayesian rule: the two probabilities $p(H) \times p(D|H)$ and $p(\neg H) \times p(D|\neg H)$ need to be computed and compared to see which one is larger. With natural frequencies, only two numbers need to be compared: H&D and ¬H&D. The same comparison needs to be made with icon arrays. However, because the quantities expressed in icon arrays are encoded faster and without intention, icon arrays should have a speed advantage over natural frequencies.

To summarize, RTs for posterior probabilities are expected to be faster for icon arrays and natural frequencies than for conditional probabilities, and should depend on the total set size. Categorical inferences should be fastest with icon arrays, followed by natural frequencies, and the slowest RTs are expected with conditional probabilities. The first part of this chapter will focus on differences in performance and RTs between the three different formats.

However, the computations required in Equations (1) and (2) do not only differ by the number of required arithmetic operations, but also by their nature. Natural frequencies only demand one addition, if the posterior probability estimate is assessed as a fraction instead of a percentage or probability, as was done in previous studies on natural frequencies in Bayesian reasoning (e.g., Gigerenzer & Hoffrage, 1995; Hoffrage et al., 2002, 2000). Conditional probabilities require subjects to multiply probabilities to determine the posterior probability. These different arithmetic operations – additions and multiplications – have been found to recruit different working-memory systems as will be delineated below. The second part of this chapter examines if Bayesian reasoning processes with different information formats involve different working-memory systems, corresponding to the arithmetic operations which are involved in categorical judgments or calculating posterior probabilities. The working-memory model of Baddeley and Hitch (1974; Baddeley, 1986) proposes a modular system that is composed of a central executive control system that supervises a verbal-auditory and a visuo-spatial slave system, which maintain information active as long as it is needed for further processing. Dehaene (1992) suggest that additions involve manipulating the position on a mental number line which requires resources of the visuo-spatial slave system. Multiplications, on the other hand, involve the retrieval of verbally stored multiplication facts, and require verbal-auditory processing capacities. Studies which used a dual-task paradigm support these assumptions. Performing a concurrent secondary task which requires capacities of the visuo-spatial system selectively impaired addition performance, but not multiplication, whereas the opposite pattern emerged for a secondary task that restricted capacities of the verbal-auditory system (Lee & Kang, 2002; Seitz & Schumann-Hengsteler, 2000). Contrary to the assumption that multiplication relies on verbal-auditory working-memory capacities because it involves a recall of verbally stored information, Imbo and van Dierendonk (2007) examined the strategies used to solve multiplication problems under dual-task conditions, and found that a verbal-auditory secondary task only interfered with multiplication performance if subjects counted the result, but not if it was recalled from memory. Multiplication performance via memory retrieval was only affected if the secondary task involved monitoring the content of short-term memory via the central executive. Therefore, conditional probabilities should primarily involve verbal-auditory components of working-memory for questions about exact posterior probabilities and categorical judgments, making performance vulnerable to interference from a verbal-auditory dual-task.

Since natural frequencies rely on additions, they should recruit visuo-spatial capacities for calculating the posterior probability. Performance should therefore deteriorate with a visuo-spatial secondary task, which should manifest in longer RTs and/or lower rates of correct solutions.

Icon arrays require counting to give posterior probabilities, which relies on verbal-auditory working-memory functions, so that slower RTs and fewer correct answers are expected when performing a verbal-auditory secondary task simultaneously.

Finally even though categorical judgments with natural frequencies and icon arrays both require only a comparison of two numbers, icon arrays should make fewer working-memory demands because they are directly processed and do not need to be encoded. In other words, Bayesian reasoning performance with icon arrays should not be adversely affected by a concurrent secondary task.

## Methods

The same experimental setup was chosen to investigate RT differences between the three formats and performance changes under dual-task conditions, described below.

### Sample

Participants were 180 university students (98 women and 82 men, mean age 26.4 years) recruited via the laboratory of the Max Planck Institute for Human Development. All participants provided written informed consent in accordance with the regulations of the Ethics Committee of the Max Planck Institute for

Human Development and received a monetary compensation contingent on their performance on the experimental tasks.

**Design and procedure**

Each participant was tested individually in a quiet environment for approximately 1 hour. Participants were randomly allocated to one of the six experimental groups derived by a full factorial design of the two factors "information format" (conditional probabilities vs. natural frequencies vs. icon arrays) and "dual-task modality" (verbal secondary task vs. visual secondary task) and answered demographic questions. Subjects then received written information about Bayesian reasoning tailored to their information format and a sample problem with the correct solutions. Subjects in the conditional probabilities group were explained Bayes rule and how to combine information, including all necessary arithmetic steps. Subjects with natural frequencies and icon arrays were instructed how to interpret the information and derive posterior probabilities and categorical inferences. Reasoning problems and dual-tasks were presented on Intel powered computers using E-Prime® software with subjects seated approximately 40 cm from the screen. All subjects solved five problems of the Bayesian reasoning task under single-task condition, and five problems while performing a concurrent secondary task (dual-task condition). In each experimental group, half of the participants started with the dual-task condition and solved the Bayesian reasoning problems without a secondary task afterwards, for the other half of participants, the order was reversed. The monetary compensation consisted of a flat fee of 9€ with the possibility to earn 10 cents for each correctly answered Bayesian reasoning problem and 10 cents if all trials of the secondary task were answered correctly while solving each reasoning problem, yielding a possible maximum of 12€.

*Bayesian reasoning task*

Ten Bayesian reasoning problems with information either as conditional probabilities, natural frequencies, or icon arrays were constructed. Each problem contained a general introduction located at the top of the screen, which introduced the general setting of the problem, a field which contained the numerical information of the problem located in the middle, and a section with a question text and a visual reminder of the answering key mappings at the bottom of the screen (see Figure 12). For each problem, subjects were asked two questions. First, they were asked if they would expect more cases conforming to the hypothesis or cases not conforming to the hypothesis stated in the question text. The problems were constructed so that both answers were correct equally often in order to eliminate the potential influence of non-task-related response preferences. Second, subjects were asked for the exact posterior probability. In the conditional probability condition, the posterior probability was queried as percentage. In the icon array and natural frequency version, subjects had to indicate a ratio which corresponds to the posterior probability. As each problem started, participants were asked to read the introduction and the question carefully and press a response button to continue. For problems in the dual-task condition, the secondary task started, and the numerical information appeared after four trials of the secondary task (6 sec). In the single-task condition, the numerical information appeared six seconds after the button press. Subjects were instructed to press one of two response buttons as soon as they determined the answer. After a pause, the general introduction reappeared, and the question about the exact posterior probability was displayed. After confirming, the numerical information appeared in the same manner as before, i.e. after four trials of the secondary task or a pause of equal length. Subjects were again instructed to press a response button to answer the question. They then had 5 seconds to enter and confirm their answer using the number pad of the keyboard. Answers given after the time expired were dismissed from the analysis to ensure all arithmetic processing was conducted under the possible influence of the secondary task.

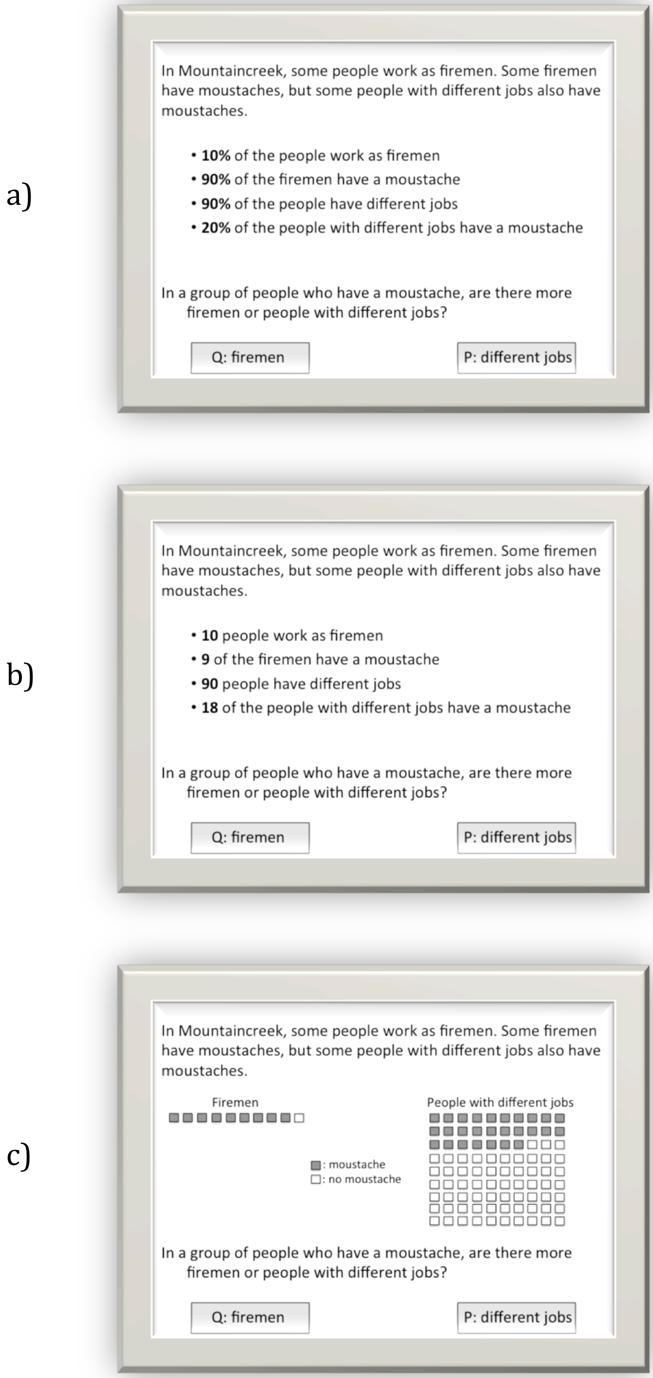



*Figure 12*. A sample problem about categorical inferences from the experiment, with different information formats: a) conditional probabilities, b) natural frequencies, and c) icon arrays. All problems contained a general introduction about the setting of the question on the top, a section with numerical information depending on the format in the middle, and a question section with answering options at the bottom of the screen.

After they submitted their answer, participants were led to a blank screen and could advance to the next problem when they were ready. At the beginning of the experiment, subjects solved three practice problems to get used to the answering mode, and feedback on the correct answers was provided. RTs and the number of correct solutions were used as dependent variables. RTs were measured from the onset of the numerical information to the time participants responded by pressing a response button. RTs less than 200 msec were treated as false reactions and discarded from the analysis.

*Secondary task*

A verbal or visual version of the *N*-back task (see Kirchner, 1958) with N = 1 was used as a secondary task. *N*-back tasks require participants to monitor a sequence of stimuli and respond if the current stimulus is identical to the one displayed *N* trials before. With *N*=1, participants have to respond if the same stimulus occurs on two successive trials. In the auditory version, four Russian nouns were used as stimuli, each 1 sec in duration, with an inter-stimulus interval (ISI) of 500 msec. Subjects had to respond if the same stimulus appeared on consecutive trials. It was ensured that the participants did not have knowledge of Russian, so that the task would involve verbal, but not semantic processing. Stimuli were presented via headphones, and participants could adjust the volume themselves. In the visual version, a black diamond with the size of 264 × 264 pixels (9.31 × 9.31 cm) appeared on the outside of one of the four corners of the numerical information field. Similar to the auditory version, stimuli were displayed for 1 sec, and the ISI was set to 500 msec. The participants were instructed to respond if the stimulus appeared at the same location on consecutive trials.

In both versions, responses were classified as correct if made after 200 msec, but before the new stimulus appeared. Prior to the experiment, participants were trained in the secondary task until they responded correctly to four successive target configurations. An index of secondary task performance was calculated by subtracting the rate of false-positive reactions from the rate of correct responses to target configurations.

## Results

To investigate differences in Bayesian reasoning between the three formats, RT and performance data under single task conditions for all three formats were pooled and analyzed together. As the order of task administration did not affect RTs and rates of correct solutions, the data of subjects starting under single task conditions and subjects starting under dual-task conditions was combined for this analysis to achieve greater power.

### Response times for Bayesian reasoning with different formats under single-task conditions

With only two exceptions, dependent variables were not normally distributed as revealed by a Kolmogorov-Smirnov test, and appropriate nonparametric methods were therefore used for data analysis.

Table 5 shows medians and ranges for correct answers to categorical and exact judgments and RTs for the three information formats under single-task conditions, both for all answers and correct answers only. A Kruskall-Wallis test confirmed that the statistical format significantly influenced the number of correct categorical judgments ($\chi^2(2, N = 180) = 12.89$, $p = .002$, $\phi = 0.27$) and posterior probabilities ($\chi^2(2, N = 180) = 48.95$, $p < .001$, $\phi = 0.52$), as well as their respective RTs (categorical judgments: all problems: $\chi^2(2, N = 180) = 66.45$, $p < .001$, $\phi = 0.61$, correct answers only: $\chi^2(2, N = 180) = 66.42$, $p < .001$, $\phi = 0.61$; exact judgments: all problems: $\chi^2(2, N = 180) = 56.60$, $p < .001$, $\phi = 0.56$, correct answers only: $\chi^2(2, N = 125) = 20.99$, $p < .001$, $\phi = 0.34$). We performed Mann-Whitney-U tests to examine differences between individual formats.

*Table 5*. Medians and ranges for correct solutions and RTs for categorical judgments and posterior probability judgments with icon arrays, natural frequencies, and conditional probabilities under single-task condtions.

| | Correct Answers | | RT[1] (All Problems) | | RT (Correct Answers Only) | |
|---|---|---|---|---|---|---|
| | Mdn[2] | Range | Mdn | Range | Mdn | Range |
| *Icon Arrays* | | | | | | |
| Categorical Judgments | 5 | 2 – 5 | 5.4 | 1.6 – 36.2 | 5.7 | 1.6 – 24.3 |
| Posterior Probabilities | 3 | 0 – 5 | 10.7 | 3.1 – 39.5 | 12.1 | 3.1 – 39.5 |
| *Natural Frequencies* | | | | | | |
| Categorical Judgments | 5 | 1 – 5 | 11.4 | 1.6 – 39.5 | 11.0 | 1.58 – 39.5 |
| Posterior Probabilities | 3 | 0 – 5 | 9.2 | 1.6 – 34.9 | 9.1 | 3.1 – 34.9 |
| *Conditional Probabilities* | | | | | | |
| Categorical Judgments | 4 | 2 – 5 | 20.5 | 3.1 – 74.4 | 19.8 | 3.1 – 74.4 |
| Posterior Probabilities | 0 | 0 – 4 | 25.1 | 4.6 – 147.1 | 28.1 | 4.6 – 147.2 |

[1] seconds

[2] median

*Categorical judgments*

Participants gave more correct answers to categorical judgments questions with icon arrays compared to conditional probabilities (z = -3.60, $p$ < .001, r = -0.33), but only small differences emerged between natural frequencies and icon arrays (z = -1.74, $p$ = .082, r = -0.16) or conditional probabilities (z = -1.18, $p$ = .070, r = -0.17). Categorical judgments with icon arrays were considerably faster than with conditional probabilities, both for all (z = -7.56, $p$ < .001, r = -0.69) and for correct answers only (z = -7.58, $p$ < .001, r = -0.71). Compared to natural frequencies, only RTs for correct answers were faster with icon arrays (z = -4.43, $p$ < .001, r = -0.41). Natural frequencies also yielded faster RTs for all (z = -4.80, $p$ < .001, r = -0.44) and correct answers (z = -4.82, $p$ < .001, r = -0.44) compared to conditional probabilities (see Figures 13 and 14).

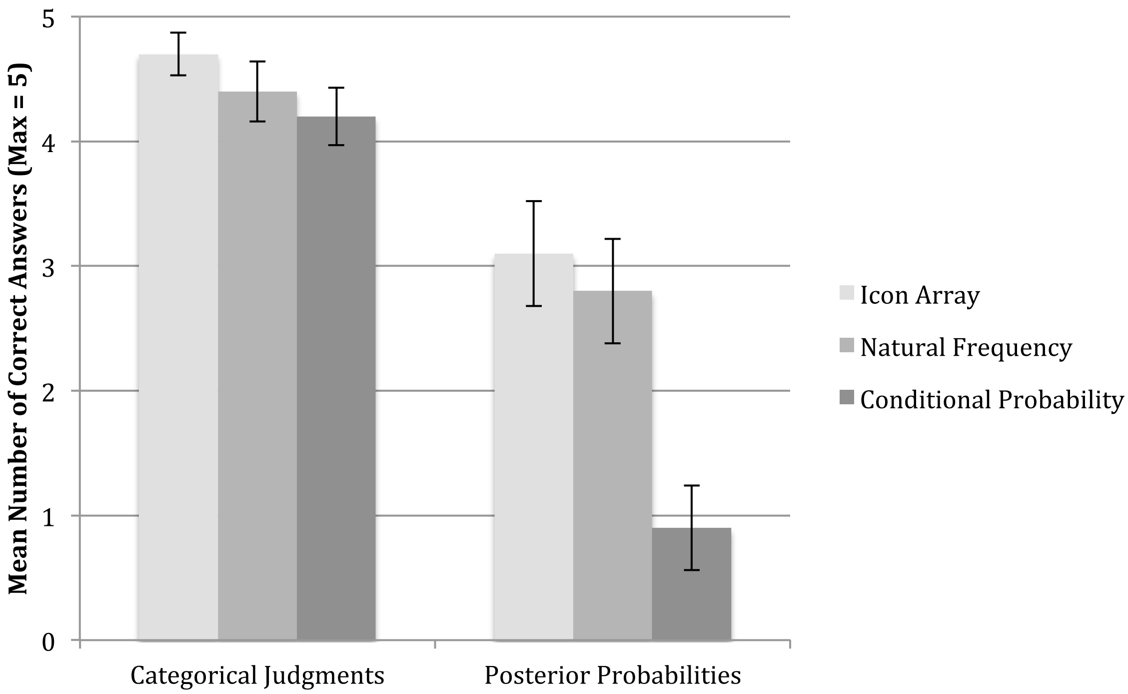


*Figure 13*. Mean number of correct categorical judgments and judgments of posterior probability observed with different information formats under single-task conditions. Error bars indicate 95% confidence intervals around the means.

*Posterior probabilities*

As can be seen in Figure 13, icon arrays and natural frequencies both led to more correct responses compared to conditional probabilities, and also yielded significantly shorter RTs, both for all problems and correctly answered problems (see Figure 14). Icon arrays and natural frequencies did not differ on the rate of correct solutions or RTs (all $z < -1.28$, *p*s $> .201$, rs $< -0.12$).

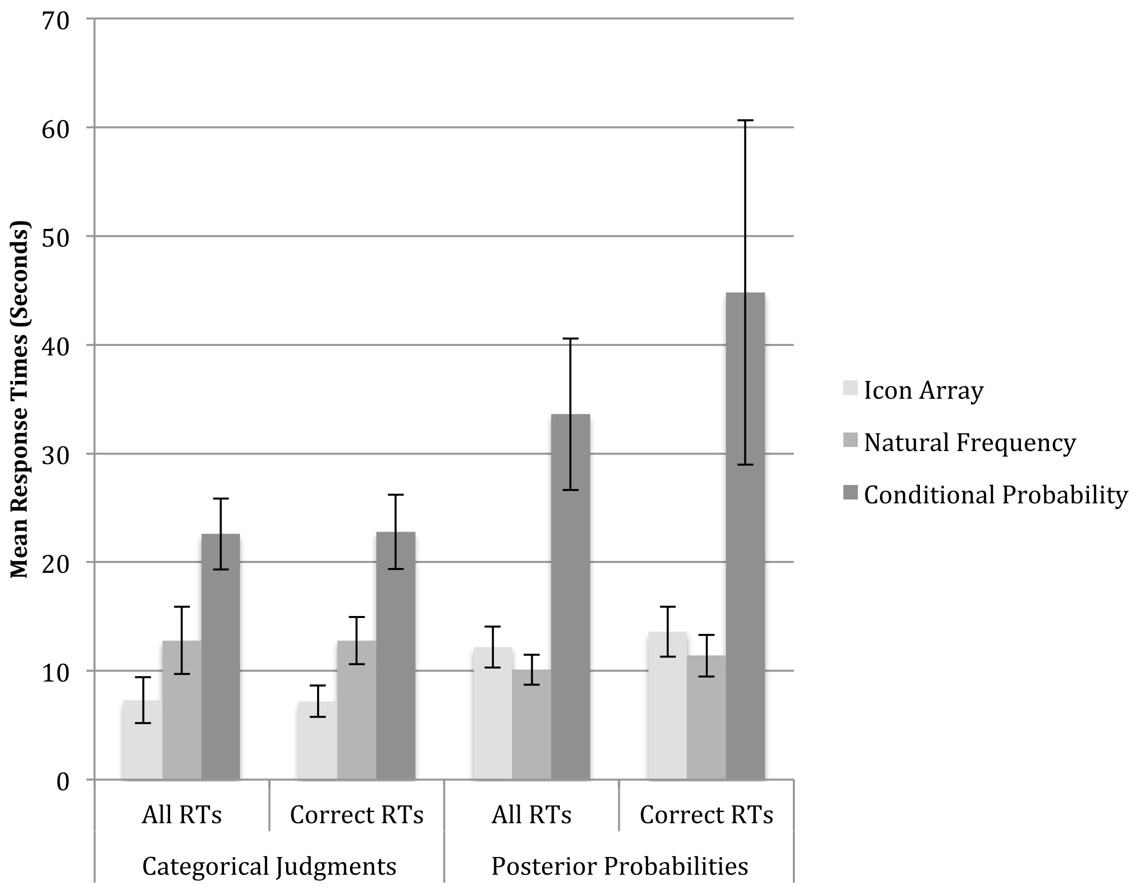



*Figure 14*. Mean response times (RTs) for categorical judgments and estimates of posterior probabilities in the icon array, natural frequency, and conditional probability condition. Error bars represent 95% confidence intervals around the mean.

*Comparison of categorical judgments and posterior probabilities*

Using a Wilcoxon signed-rank test, we compared the number of correct solutions and RTs for categorical inferences to those for posterior probabilities. For all information formats, we observed a higher number of correct solutions for categorical inferences than for posterior probabilities. RTs were higher for posterior probabilities compared to categorical judgments for icon arrays and conditional probabilities, but yielded only small effects for natural frequencies (all answers: z = -1.94, $p$ = .053, r = -.25; correct answers only: z = -1.09, $p$ = .275, r = -.16).

*Correlations*

For the icon array and natural frequencies condition, we determined the total size of the set which had to be considered in order to give the correct exact posterior probability. The total set-size consists of the denominator of Bayes rule, as shown in Equation (2), i.e. H&D + ¬H&D. For the problems used, the set-sizes varied between 10 and 189.

In the icon array condition, the RT for exact posterior probabilities was correlated with the total numbers of icons which had to be counted, both for all responses ($r_s = 0.34$, $p < .001$) and correct responses only ($r_s = 0.35$, $p < .001$). For natural frequencies, we also observed a correlation between the set-size and RTs to posterior probabilities for all answers ($r_s = 0.18$, $p = .002$) and correct answers only ($r_s = 0.28$, $p < .001$).

**Bayesian reasoning performance under dual-task conditions**

To analyze the impact of visual or auditory dual-task performance on Bayesian inferences, each participant's performance under single task conditions was compared to performance under dual-task conditions. All participants solved the Bayesian reasoning problems both under single- and dual-task conditions. Therefore a Wilcoxon rank sum test was used to compare differences in the number of correct solutions and RTs for categorical inferences and posterior probability judgments. Since performance of participants who started under single- or dual-task conditions did not differ, the data was combined for the analysis.

*Categorical judgments*

Table 6 contains information on descriptive parameters of categorical judgments. As the effects of dual- vs. single-task performance were often not reflected in the medians, Figure 15 shows the number of participants who solved more or fewer problems correctly under dual-task compared to single-task conditions for both secondary task modalities. With icon arrays, participants in the verbal dual-task condition answered more problems correctly under dual-task conditions than when solving the Bayesian problems alone (z = -2.33, $p$ = .020, r = -.43), while RTs did not differ (all RTs: z = -1.29, $p$ = .198, r = -.23; correct answers only: z = -1.20, $p$ = .232, r = -.22). For the visual dual-task condition, we observed a similar, but smaller and non-significant effect for the number of correct solutions (z = -1.51, $p$ = .13, r = -.28). Subjects showed higher RTs compared to the single-task condition (z = -2.12, $p$ = .034, r = -.39), but only a small and insignificant effect emerged analyzing only RTs for correct answers (z = -1.30, $p$ = .194, r = -.24).

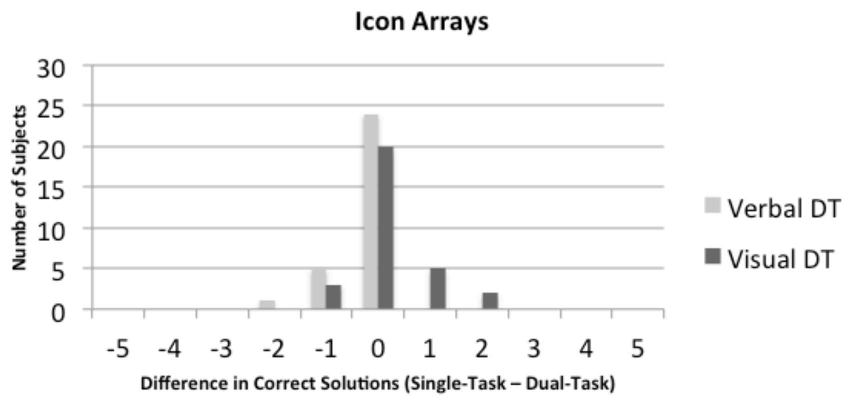
For the natural frequency condition, the number of correct responses under single- and dual-task conditions did not differ for both secondary task modalities (auditory secondary task: z = -0.93, $p$ = .355, r = -.17; visual secondary task: z = -0.73, $p$ = .465, r = -.13). However, subjects in both secondary task conditions showed higher RTs for all problems (auditory dual-task: z = -3.08, $p$ = .002, r = -.56; visual dual-task: z = -2.90, $p$ = .004, r = -.52) and correctly answered problems (auditory dual-task: z = -2.81, $p$ = .005, r = -.52; visual dual-task: z = -2.78, $p$ = .005, r = -.51).

With conditional probabilities, subjects answered fewer questions correctly in the auditory (z = -1.92, $p$ = .055, r = -.35), but not in the visual dual-task condition (z = -0.35, $p$ = .729, r = -.06). RTs were slower under both dual-task conditions, both for all problems (auditory dual-task: z = -2.32, $p$ = .020, r = -.42; visual dual-task: z = -2.57, $p$ = .010, r = -.47) and correct answers only (auditory dual-task: z = -2.09, $p$ = .037, r = -.38; visual dual-task: z = -1.78, $p$ = .074, r = -.33).

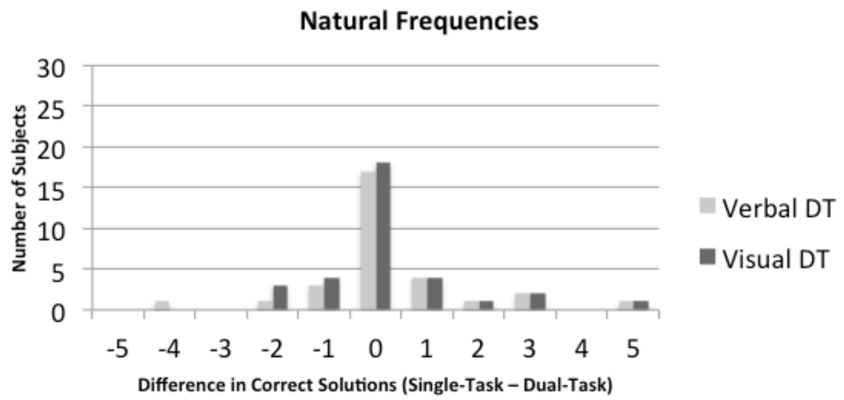
*Table 6*. Medians and ranges for the number of correct solutions and RTs for all answers and correct answers only for categorical inferences. Data for both dual-task (DT) modalities is reported separately.

| | Single-Task Condition | | | | | | Dual-Task Condition | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Correct Answers | | RT (all)[1] | | RT (correct only) | | Correct Answers | | RT (all) | | RT (correct only) | |
| | Mdn[2] | Range | Mdn | Range | Mdn | Range | Mdn | Range | Mdn | Range | Mdn | Range |
| *Icon Arrays* | | | | | | | | | | | | |
| Visual DT[3] | 5 | 3 - 5 | 4.6 | 1.6 – 24.3 | 4.6 | 1.6 – 24.3 | 5 | 3 - 5 | 6.1 | 1.6 – 24.3 | 5.0 | 1.6 – 24.3 |
| Verbal DT | 5 | 2 - 5 | 6.1 | 1.6 – 36.2 | 6.1 | 1.6 – 23.4 | 5 | 2 - 5 | 6.1 | 1.6 – 31.8 | 6.5 | 1.6 – 31.8 |
| *Natural Frequencies* | | | | | | | | | | | | |
| Visual DT | 5 | 2 - 5 | 12.2 | 4.6 – 31.9 | 12.2 | 4.6 – 31.9 | 5 | 0 - 5 | 16.7 | 4.6 – 43.9 | 16.7 | 4.6 – 43.9 |
| Verbal DT | 5 | 1 - 5 | 9.9 | 1.6 – 39.5 | 9.9 | 1.6 – 39.5 | 5 | 0 - 5 | 15.2 | 3.1 – 62.2 | 15.2 | 3.1 – 60.0 |
| *Conditional Probabilities* | | | | | | | | | | | | |
| Visual DT | 4 | 2 - 5 | 19.8 | 3.1 – 62.1 | 19.4 | 3.1 – 62.1 | 5 | 1 - 5 | 26.5 | 4.6 – 54.7 | 24.3 | 6.1 – 52.4 |
| Verbal DT | 4.5 | 2 - 5 | 22.8 | 4.6 – 74.4 | 21.3 | 6.9 – 74.4 | 4 | 1 - 5 | 2.,3 | 4.6 – 150.6 | 25.8 | 4.6 – 150.6 |

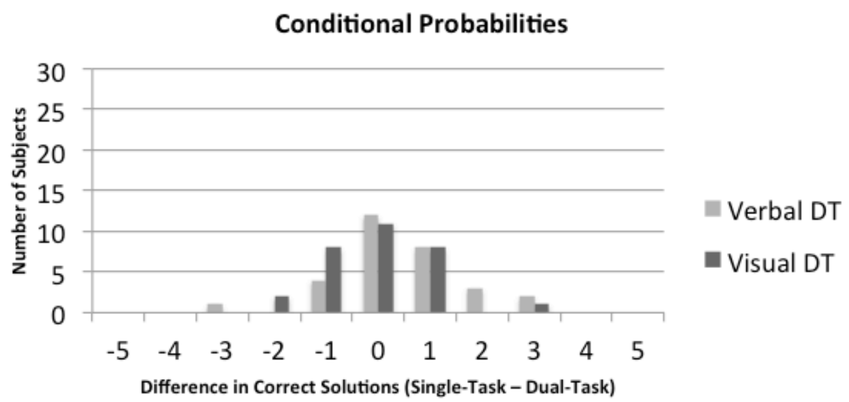
[1] seconds
[2] median
[3] dual-task




a)




b)




c)

*Figure 15.* Modality-specific differences between the number of correct categorical judgments under single- and dual-task conditions for (a) icon arrays, (b) natural frequencies, and (c) conditional probabilities. The figure shows the number of participants in each experimental condition who solved more, less, or an equal number of questions correctly under single- and dual-task conditions. Positive numbers indicate that more problems were solved under single-task conditions than under dual-task conditions. E.g., 5 subjects in the verbal dual-task condition (verbal DT) with icon arrays

answered one more problem correctly under dual-task conditions compared to single-task performance, and two subjects answered even gave two more correct answers.

*Posterior probability judgments*

Subject-level comparisons of single- and dual-task performance for judgments of posterior probabilities can be found in Figure 16. With icon arrays, the number of correct answers did not differ under single- and dual-task conditions for both modalities (auditory dual-task: z = -0.80 *p* = .424, r = -.15; visual dual-task: z = -0.63, *p* = .539, r = -.12). However, RTs were higher with the auditory dual-task (all problems: z = -2.95, *p* = .003, r = -.54; correct answers only: z = -1.98, *p* = .050, r = -.40), but not with the visual dual-task (all problems: z = -1.45, *p* = .147, r = -.26; correct answers only: z = -0.75, *p* = .454, r = -.15).

For subjects with natural frequency information, we also did not find an effect of dual-task performance on the number of correct solutions (auditory dual-task: z = -0.93 *p* = .355, r = -.17; visual dual-task: z = -0.56, *p* = .58, r = -.10). RTs were only affected by the visual dual-task (all problems: z = -2.16, *p* = .031, r = -.39; correct answers only: z = -1.83, *p* = .067, r = -.38), but not the auditory dual-task (all problems: z = -1.62, *p* = .105, r = -.30; correct answers only: z = -0.71, *p* = .475, r = -.15).

Participants in the conditional probability condition answered fewer problems correctly with an auditory dual-task (z = -1.93, *p* = .054, r = -.35), but not with a visual dual-task (z = -0.09, *p* = .933, r = -.02). With the auditory dual-task, RTs were not different from single-task performance (all problems: z = -1.33, *p* = .183, r = .24; correct answers only: z = -1.83, *p* = .068, r = -.82). The same result emerged for the visual dual-task condition (all problems: z = -.141, *p* = .888, r = -

.03; correct answers only: z = -1.27, $p$ = .203, r = -.40). See Table 7 for descriptive information about posterior probabilities.

**Secondary task performance**

Secondary task performance was analyzed to ensure that the observed effects on Bayesian reasoning performance were not caused by a shift of attentional resources from the reasoning task to the secondary task, and that conversely other possible effects were not concealed by detracting resources from the secondary task. A Kolmogorov-Smirnov test revealed that secondary task accuracy was normally distributed within the groups. Therefore, we performed a random-effects 3×2 analysis of variance with "information format" and "dual-task modality" as between-subjects factors. Accuracy values varied between 0.63 and 0.72, with standard deviations between 0.15 and 0.21. The analysis revealed neither a significant influence of "information format" ($F(2, 169) = 0.69$, $p = .590$, $\eta^2 = 0.41$), "dual-task modality" ($F(1, 169) = 0.276$, $p = .651$, $\eta^2 = 0.12$), nor an interaction effect ($F(2, 169) = 1.193$, $p = .306$, $\eta^2 = 0.01$).

*Table 7*. Medians and ranges for the number of correct solutions and RTs for all answers and correct answers only for posterior probabilities.

| | Single-Task Condition | | | | | | Dual-Task Condition | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Correct Answers | | RT (all)[1] | | RT (correct only) | | Correct Answers | | RT (all) | | RT (correct only) | |
| | Mdn[2] | Range | Mdn | Range | Mdn | Range | Mdn | Range | Mdn | Range | Mdn | Range |
| *Icon Arrays* | | | | | | | | | | | | |
| Visual DT[3] | 4 | 0 - 5 | 10.6 | 3.1 – 39.4 | 11.4 | 3.1 – 39.4 | 3.5 | 0 – 5 | 11.4 | 4.6 – 53.1 | 11.4 | 4.6 – 53.1 |
| Verbal DT | 3 | 0 - 5 | 12.1 | 4.6 – 24.3 | 12.1 | 4.6 – 24.3 | 3 | 0 – 5 | 15.2 | 3.1 – 72.8 | 15.2 | 3.1 – 66.6 |
| *Natural Frequencies* | | | | | | | | | | | | |
| Visual DT | 3 | 0 - 5 | 8.8 | 3.1 – 33.4 | 8.0 | 3.1 – 34.9 | 3.5 | 0 – 5 | 12.2 | 2.3 – 47. 5 | 9.1 | 2.3 – 47.1 |
| Verbal DT | 3 | 0 - 5 | 9.1 | 1.6 – 18.2 | 10.7 | 4.6 – 22.8 | 3.5 | 0 – 5 | 12.2 | 1.6 – 59.0 | 12.9 | 1.6 – 59.0 |
| *Conditional Probabilities* | | | | | | | | | | | | |
| Visual DT | 0.5 | 0 - 4 | 22.0 | 4.6 – 147.1 | 19.8 | 4.6 – 147.1 | 1 | 0 – 4 | 25 | 2.3 – 72.7 | 23.1 | 3.1 – 47.1 |
| Verbal DT | 0 | 0 - 5 | 27.3 | 6.1 – 105.9 | 28.8 | 16.7 – 105.9 | 0 | 0 – 3 | 24.3 | 1.6 – 153.2 | 8.4 | 1.6 – 42.4 |

[1] seconds
[2] median
[3] dual-task

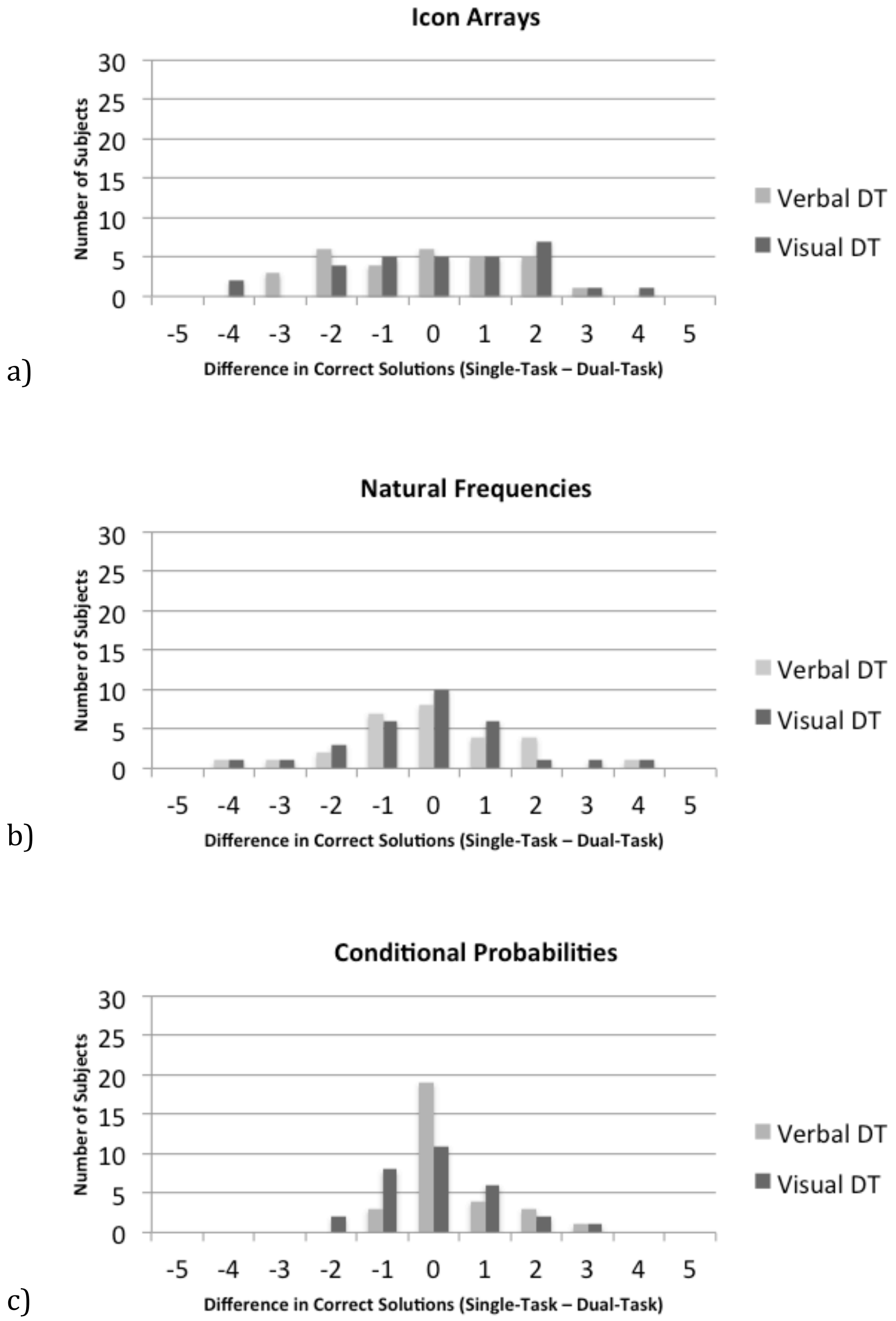



*Figure 16*. Differences between the number of correct judgments of posterior probability under single- and dual-task conditions for (a) icon arrays, (b) natural frequencies, and (c) conditional probabilities. The figure shows the number of participants in each experimental condition who solved more, less, or an equal number of questions about posterior probabilities correctly under single- and dual-task conditions. Positive numbers indicate that more problems were solved under single-task conditions than under dual-task conditions.

## Discussion

This chapter examined how different arithmetic and mental operations induced by different information and question formats are linked to rates of correct responses and RTs. The first part of this analysis has focused on using RTs as additional indicators of cognitive load. The second part has shown that the information and response format involve different arithmetic operations, and ultimately rely on different components of working-memory.

### Response time differences in Bayesian reasoning with different information formants

The results confirm previous findings about the advantage of natural frequencies and icon arrays over conditional probability information in producing more correct judgments of posterior probability (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Hoffrage et al., 2002). We did not, however, find an advantage of icon arrays over natural frequencies, as reported by Cosmides and Tooby (1996). In their experiment, an icon array had to be actively constructed from natural frequency information, and was found to produce correct solutions for 92% of their subjects, whereas natural frequencies elicited correct responses in 76% of the participants. They also tested another condition where subjects were shown an already constructed icon array, and rates of correct solutions matched those of natural frequencies. In our study, icon arrays did not need to be constructed by our participants, but were already provided. Galesic et al. (2009) found that icon arrays only provide an advantage over natural frequencies for people with low numerical abilities. Participants in our study were university students, who likely possess above-average numerical abilities and are therefore less likely to benefit from icon arrays.

This study for the first time reported RT data for Bayesian inferences. For questions about posterior probabilities, the information format had a similar impact on RTs as on the number of correct answers. With icon array or natural frequency information, subjects could derive correct answers more often and faster, compared to conditional probability information. These results are in accordance with the claim that computational facilitation drives the effect between the formats. Icon arrays and natural frequencies both require the same number of arithmetic operations, but fewer than conditional probabilities. Each additional arithmetic operation is a chance for calculation errors and produces an increase in RTs (Ashcraft, 1992). As expected from research in mental arithmetic, we observed a correlation of RTs for natural frequency information with the size of the set that was used in a problem, corresponding to the set-size effect. The same relationship was found for icon arrays. Here, participants had to count the individual icons to determine H&D and ¬H&D to calculate the correct posterior probability correctly. Since counting is a serial operation, RTs increase with each icon in the set that needs to be counted.

Categorical inferences were easier to perform than calculating posterior probabilities. The number of correct responses was highest for icon arrays, whereas the difference between natural frequencies and conditional probabilities was rather small. Also, RTs were highest with conditional probabilities and lowest for icon arrays. The faster RTs for correct categorical judgments with icon arrays as opposed to natural frequencies could reflect the difference in quantity encoding. Icon arrays display quantities as single, countable objects which are encoded without intention, whereas the association of Arabic numerals like in natural frequencies with an analogue representation of quantity is established later in development, which may cause a processing advantage reflected in shorter RTs. Overall, subjects showed higher rates of correct categorical inferences compared to posterior probabilities, and RTs which reflected the processing demands.

In accordance with computational facilitation explanations, RT and performance data for Bayesian inference tasks go in the same direction: more computational steps were associated with longer RTs and more errors. In agreement with the claim that displaying information in a way that humans have been most accustomed to throughout evolutionary development, icon arrays were processed fastest when the response format did not involve a numerical output. Using categorical judgments as an indication of Bayesian information processing resulted in a less pessimistic view of human inference abilities. Categorical judgments were solved correctly more often, and the differences between the information formats were smaller.

**Working-memory load in Bayesian reasoning with different information formats**

The second aim of this chapter was to analyze the involvement of visuo-spatial and verbal-auditory working-memory components in Bayesian reasoning tasks, depending on the arithmetic operations implied by the task and information format. The results confirm that information format and task demands determine which aspects of working-memory become involved in Bayesian reasoning.

*Information format*

The arithmetic operations implied by presenting information as conditional probabilities, natural frequencies, or icon arrays resulted in the expected pattern of interference from a secondary task. In the conditional probability format, fewer correct answers could be given when the secondary task required verbal-auditory capacities, but not for a visual secondary task. This result is consistent with the hypothesis that conditional probabilities rely on performing multiplications, and previous results which have linked multiplications to verbal-auditory working-memory functions (Imbo & Vandierendonck, 2007; Seitz & Schumann-Hengsteler, 2000). We also observed a trend for participants in the verbal dual-task condition to show much faster RTs for correct responses than under single-task conditions (medians: 8.39 sec vs. 28.06 sec). Although this effect seems surprising, the RT data originates from only 5 subjects who solved one or more of the posterior probability questions correctly. In a pilot to the experiment, some subjects in the verbal dual-task condition indicated that they memorized the information when they solved the categorical judgments to avoid having to perform the secondary task and Bayesian reasoning operations simultaneously. The few short RTs observed in this condition hint toward the same strategy, which would strengthen the conclusion that in particular verbal secondary tasks interfere with Bayesian reasoning operations with conditional probabilities, so that subjects refrain from performing them simultaneously.

Natural frequencies, on the other hand, only require one addition, which is reflected in slower RTs with a visual, but not a verbal secondary task. Finally, calculating the correct posterior probability with icon arrays requires counting the icons to give an exact answer. RTs were slower with a verbal secondary task, but not with a visual task, which is in line with findings that counting requires resources from verbal-auditory working-memory (Ketelsen & Welsh, 2010).

*Task demands*

Assessing Bayesian reasoning performance via categorical judgments or judgments on the posterior probability demands different operations, depending on the information format. With conditional probabilities, both questions rely on multiplication: subjects have to calculate the two probabilities which constitute the denominator of Bayes rule – $p(H) \times p(D|H)$ and $p(\neg H) \times p(D|\neg H)$ – and compare them to find out which is higher. Correspondingly, we found that both, categorical judgments and judgments of posterior probability, recruit resources from verbal-auditory working-memory.

Natural frequencies only demand a simple comparison between two numbers to infer if cases confirming to or contradicting the hypothesis would be more common, while calculating the posterior probability requires one addition. Consistently, judgments of posterior probabilities relied on visuo-spatial working-memory resources, while categorical judgments were susceptible to visual and verbal interference.

With icon arrays, making categorical inferences only requires a comparison of two visually presented quantities, and performance under both dual-tasks did not decline. However, icons have to be counted to give posterior probability judgments, and in agreement with studies which have linked counting to verbal working-memory, we observed longer RTs under auditory dual-task conditions, but not for a visual secondary task.

**Computational facilitation and ecological design**

Two explanations have been proposed for higher rates of correct solutions with natural frequencies and icon arrays compared to conditional probabilities: the ecological format of information which mimics how information is encountered in natural environments, and the computational facilitation which goes hand in hand with the more ecological representation, first described by Gigerenzer & Hoffrage (1995). Our results support both claims. Comparing natural frequencies to conditional probabilities, our data show that both formats do not only differ in the number of computational steps, but also in the nature of those operations, which leads to a differential involvement of working-memory systems. A visual secondary task affected Bayesian reasoning operations with natural frequencies, but not with conditional probabilities, whereas we observed the opposite pattern for a verbal secondary task.

Using ecological information formats like icon arrays decreased working-memory demands for Bayesian reasoning operations, as indicated by categorical judgments. While categorical judgments with natural frequencies interfered with both secondary tasks, categorical judgments with icon arrays were not negatively affected, to the contrary: subjects gave even more correct answers when they performed a verbal secondary task at the same time, without showing longer RTs. Under the assumption that allocating more attentional resources to a task improves performance, this pattern is unexpected. However, whether devoting more attentional resources lead to an improvement of performance seems to be mediated by the level of experience with the task: Beilock, Bertenthal, McCoy and Carr (2004; Beilock, Carr, MacMahon, & Starkes, 2002) examined putting accuracy for expert and novice golfers under time constraints and with or without a secondary task. While novice golfers were less accurate when they had fewer time to think about their actions or had to perform a secondary task, experts' putting accuracy improved with time constraints and under dual-task conditions. In novices, the motor actions required for putting are not yet fully automatized, and focusing on the movements can improve performance. Experts, however, rely on highly automatized routines, and increasing attention to some elements of the routine may disrupt performance. It is possible that the same effect occurred for icon arrays and natural frequencies. Even children can extract quantity information from a set of objects presented to them (Xu & Spelke, 2000). When seeing a group of dots, children in preschool process quantity information automatically, i.e. without deliberate effort, while this happens later for Arabic number symbols (Gebuis et al., 2009). Possibly, subjects with icon arrays may have thought too much about how to interpret the information or the question and were diverted from the correct answer. However, this explanation requires further scrutiny and should be subjected to replication.

The results presented in this chapter could inform the design of decision aids in medical settings. Doctors and patients need to understand medical information like test results, and they often have only limited time and resources. In one study reported by Gigerenzer et al. (2007), 180 Gynecologists were given information about the base rate of breast cancer and the sensitivity and false-positive rate of mammography screenings, and were asked to estimate the probability that a woman who receives a positive result in a mammography screening actually has breast cancer. While the best estimate is around 10%, most of the gynecologist believed it to be 81% or higher, vastly overestimating the predictive accuracy of mammography screenings. After the gynecologists learned to convert conditional probabilities into natural frequencies, 87% of the doctors could give the correct estimate. False-positive results in mammography screening have been shown to increase anxiety and worries about breast cancer in women (Brewer, Salz, & Lillie, 2007). Our results suggest that natural frequencies do not only improve understanding, but are also quick and resistant against interferences from concurrent tasks – and may therefore be an ideal tool for doctors and patients who have only limited time and resources available.

# V – Conclusion

This thesis began with the proposition that Bayesian reasoning is not intricately difficult for humans, but that different representations of Bayesian reasoning problems require more or less experience with symbolic numerical representations, impose more or less complicated arithmetic processing, and rely more or less on sparse attentional resources. The research presented in this dissertation confirmed this assertion and has raised new questions which may serve as starting points for future endeavors.

As chapter II demonstrated, already in second grade children can follow a Bayesian reasoning strategy systematically and give precise estimates of posterior probabilities when information is presented as single, countable objects. This is especially remarkable since children's behavior in this study was spontaneous, i.e., children did not receive a prior training in probability calculus. The approach chosen in this study may show how to teach elementary principles of statistical reasoning even to children at the beginning of formal education, a skill lacking in many parts of today's society and becoming even more important in the future. "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write" is a prediction commonly attributed to utopist writer H. G. Wells. Teaching statistical thinking early is possible, and may be the best chance to vaccinate children against statistical illiteracy that today has clouded the minds of politicians, doctors, patients, journalists and lay people (cf. Gigerenzer et al., 2007). Mathematics education is beginning to realize the need to teach children how to cope with uncertainties and risks in our daily lives (see Bond, 2009). Simply explaining how to use Bayes rule leads to only small and temporary increases in Bayesian thinking, whereas teaching how to change the representation of the problem yields large and long-lasting effects (Sedlmeier & Gigerenzer, 2001). This representational change is at the core of first programs which attempt to teach children in elementary schools basic probabilistic concepts such as Bayesian reasoning or the information content of a question by using tools such as tinker cubes or icon arrays (Kurz-Milcke, Gigerenzer, & Martignon, 2008; Latten, Martignon, Monti, & Multmeier, 2011; Martignon & Krauss, 2009). The results presented in chapter II provide a theoretical foundation for those programs and could guide future development.

The results presented in chapter III extend those of chapter II to a different population: children with developmental dyscalculia. Remarkably, those children showed a level of Bayesian reasoning performance similar to that of healthy children in chapter II when the information was presented as analogue objects. This effect, however, depends on the etiology of dyscalculia and the extent to which children developed analogue quantity representations. Training programs targeted at restoring and refining analogue representations of quantity (e.g., Wilson, Dehaene, et al., 2006; Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006) showed a marked increase in arithmetic proficiency. But even after extensive training, deficits are rarely completely remediated, and are carried on into adulthood. MacDougall (2009), for example, identifies dyscalculia as one of the largely unrecognized conditions which may affect the performance of undergraduate medical students. Those students may be left with a poor sense for important health care information, and may even not be reached by transparent statistical formats like natural frequencies. Using analogue material like icon arrays, health care information can be dispensed to and appreciated by a wider audience, especially for people with a poor understanding of basic probabilistic facts (Galesic et al., 2009).

Finally, chapter IV showed that the computational facilitation provided by icon arrays and natural frequencies is not only reflected in the rates of correct estimates of posterior probabilities, but also in RTs. Future studies interested in the mental processes behind Bayesian reasoning operations could investigate if RT analyses can also be used to differentiate Bayesian reasoning strategies. Also, while processing load naturally varied *between* information formats, subsequent research could manipulate the complexity of arithmetic processing *within* the information formats. For example, conditional probabilities could be tested against conjunct probabilities, i.e. providing subjects with the products of $p(H) \times p(D|H)$ and $p(\neg H) \times p(D|\neg H)$, thereby eliminating multiplications from Bayes rule.

The second part of chapter IV showed that icon array, natural frequencies, and conditional probabilities do not only differ in the number of computational steps, but also in the nature of the arithmetic operations, and ultimately in their working memory demands. Displaying information as icon arrays helps to gain a quick sense of the direction of the information and is particularly robust against diminishing working-memory capacities. Icon arrays may be especially helpful to convey health statistics to populations with limited working-memory capacities, e.g., elderly patients (Balota, Dolan, & Duchek, 2000; Grady & Craik, 2000; Zacks & Hasher, 2007).

Another focus of this dissertation was the level of detail Bayesian reasoning performance is assessed. Previous studies concentrated on judgments of exact posterior probabilities, but other, cruder, measures may also be indicative of Bayesian reasoning ability. In all studies reported in this dissertation, higher rates of correct solutions were observed when subjects made categorical inferences, rather than indicate precise posterior probabilities. On the one hand, this result may not be surprising because categorical judgments require fewer computational steps, and the binary forced-choice format likely overestimates true performance due to a high probability to guess the answer correctly. On the other hand, categorical judgments may be closely linked to actions: In the study of Girotto and Gonzales (2008), children had to bet on the outcome of a random draw from a bag of objects differing in shape and color. Children were asked for their initial bets, i.e. their understanding of the prior distribution of the objects, and after the experimenter hinted the shape of the drawn object, shifting the odds in favor of the color which was less frequently represented in the total sample. Children were able to detect this change in probability and changed their bet to the color now favored by Bayesian calculations. The children in their study could revise their opinions in the light of new evidence and could adapt their decisions, choosing the alternative with a higher probability of winning. Although judgments of exact posterior probabilities have not been assessed, the children in their study, much as in the studies presented in this dissertation, could indicate which alternative was favored in light of previous knowledge and new evidence. As the final section will show, choosing exact judgments of posterior probabilities as a gold standard for human information processing originates in the classical idea of rationality. This classical notion, however, has been criticized, and alternative accounts of rational behavior have been proposed, which do not require precise probability estimates for adaptive behavior.

Classical models of rational decision-making of the expected value family (Pascal, Eliot, & Trotter, 2003) like expected utility maximization (Edwards, 1954; Neumann & Morgenstern, 1953), prospect theory (Kahneman & Tversky, 1979), or cumulative prospect theory (Tversky & Kahneman, 1992) all assume that rational decision-makers weigh the value or utility of an alternative with the probability that this alternative will eventually occur and opt for the one with the highest product. Thus, not only determining the value or utility of alternatives correctly is crucial, but also estimating the probability that this option will eventually occur. When Ward Edwards, proponent of the expected utility approach, conducted his studies in the 1960s, the focus was proximately on Bayesian inferences, but ultimately the motivation was studying if people could accurately update their beliefs about the world as a precondition of rational decision-making. Throughout the transformation of classical decision-making theories, the weights given to probability estimates changed, but at the core those models still assumed that humans evaluate each option with a precise probability.

More recently, the descriptive, normative, and prescriptive nature of the classical models of decision-making has been criticized (Gigerenzer, 1991; Gigerenzer, 1996). In a new program termed the Adaptive Toolbox Approach (Gigerenzer, Todd, & the ABC Research Group, 1999), human decision-making is described by a set of heuristics, simple algorithms which exploit the information structure of the environment and allow for ecologically rational decisions. The notion of *ecological rationality* was heavily influenced by Egon Brunswik and acknowledges the fact that human cognition is adapted to the environments in which it evolved, and that cognition makes use of the information structure in natural environments for inferences and decisions. One of the heuristics, Take The Best (Gigerenzer & Goldstein, 1996), uses cues to infer which of two object scores higher on a criterion, such as city population size, male and female attractiveness, high school drop out rates, house prices, or ozone levels in San Francisco (see Czerlinski, Gigerenzer, & Goldstein, 1999). In the case of city population, cues can be whether a city has an airport, is a state capital, has a soccer team in the major German soccer league, or if it has a major university. The cues are ranked by their validities, and the heuristic then goes through the cues starting with the highest validity and stops when one cue differentiates between the two objects. At this point, the inference is made that the object with the positive cue value (e.g., city A is a state capital, city B is not) is the one higher on the criterion, and search for more information is terminated. Deciding which city has more inhabitants therefore does not require estimates of probabilities on the side of the subjects. However, Bayesian thinking is required when ranking the cues according to their validities. The validity of a cue can be described as the probability that it makes a correct inference, given that the cue differentiates between the objects and thus allows making an inference. But instead of calculating exact probabilities, Take The Best only demands a ranking of cues according to their posterior probabilities (see the Appendix in Martignon & Krauss, 2009). Judging whether being a state capital or having a soccer club in the major soccer league is more predictive for city population may therefore be a level of Bayesian reasoning that people actually need in their daily lives to make decisions. First research on the development of heuristic decision-making suggests that even children can generate and order cues just as well as, or even better than, adults (Ruggeri & Katsikopoulos, under review).

Following the enlightenment idea of probability as a “universal calculus” of the mind (Leibniz, 1677), researchers have tried to describe psychological processes in the terminology of probability calculus, emphasizing the internal consistency of beliefs. However, following the functionalist view going back to William James, human cognition is adapted to environments and specific tasks which are important to survive and reproduce in an uncertain world. Having functionality, not mathematical accuracy or logical consistency as the goal of cognition might shift the criterion against which human Bayesian reasoning performance has to be evaluated. Ecological design does not only apply to the presentation of information, but also to the task given to subjects in laboratory experiments. Just like choosing an information format which does not maintain the information structure of the environment, imposing tasks on subjects outside the realms of those important to their fitness or their everyday functioning may distort the picture of humans’ cognitive abilities. In that case, the bias resides in the minds of the researchers, rather than their subjects’.

# VI – References


Alarcon, M., DeFries, J. C., Light, J. G., & Pennington, B. F. (1997). A twin study of mathematics disability. *Journal of Learning Disabilities*, *30*(6), 617–23.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1-2), 75–106.

von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology*, *49*, 868–873.

Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4 +/- 0: a new look at visual numerosity judgements. *Perception*, *5*(3), 327–334.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Recent advances in learning and motivation (vol. 8)*. New York: Academic Press.

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 395–409). New York, NY: Oxford University Press.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*, 211–233.

Barth, H., La Mont, K., Lipton, J. S., Kanwisher, N., & Spelke, E. S. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*, 199–222.

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review*, *11*(2), 373–379.

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology. Applied*, *8*(1), 6–16.

Betsch, T., Biel, G., Eddelbüttel, C., & Mock, A. (1998). Natural sampling and base‐rate neglect. *European Journal of Social Psychology*, *28*(2), 269–273.

Boddaert, N., Mochel, F., Meresse, I., Seidenwurm, D., Cachia, A., Brunelle, F., Lyonnet, S., et al. (2006). Parieto-occipital grey matter abnormalities in children with Williams syndrome. *NeuroImage*, *30*(3), 721–725.

Bond, M. (2009). Decision-making: Risk school. *Nature*, *461*(7268), 1189–1192.

Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, *23*, 369–381.

Brewer, N. T., Salz, T., & Lillie, S. E. (2007). Systematic review: the long-term effects of false-positive mammograms. *Annals of Internal Medicine*, *146*(7), 502–510.

Bruandet, M., Molko, N., Cohen, L., & Dehaene, S. (2004). A cognitive characterization of dyscalculia in Turner syndrome. *Neuropsychologia*, *42*(3), 288–298.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459.

Cappelletti, M., Barth, H., Fregni, F., Spelke, E. S., & Pascual-Leone, A. (2007). rTMS over the intraparietal sulcus disrupts numerosity processing. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *179*(4), 631–642.

Church, R. M., & Meck, W. H. (1984). The numerical attribute of stimuli. In H. T. Roitblat, T. G. Brever, & H. S. Terrace (Eds.), *Animal cognition*. Hillsdale, New Jersey: Erlbaum.

Cole, W. G., & Davidson, J. E. (1989). Graphic Representation Can Lead To Fast and Accurate Bayesian Reasoning. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 227–231.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. (1999). How good are simple heuristics? In G Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple Heuristics That Make Us Smart* (pp. 97–118). Oxford University Press. Retrieved January 16, 2012, from http://psycnet.apa.org/psycinfo/1999-04366-005

Dehaene, S, Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396.

Dehaene, S, Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opionion in Neurobiology*, *14*, 218–224.

Dehaene, S, Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*, 487–506.

Dehaene, Stanislas. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42.

Dehaene, Stanislas. (1997). *The number sense: how the mind creates mathematics*. New York, New York: Oxford University Press.

Dehaene, Stanislas. (2001). Precis of The Number Sense. *Mind and Language*, *16*(1), 16–36.

Donders, F. C. (1868). Over de snelheid van psychische processen. Onderzoekingen gedann in het Psychologisch Laboratorium der Utrechtsche Hoogeschool: 1868-1869. *Tweede Reeks*, *2*, 92–120.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In D Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge, MA: Cambridge University Press.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380–417.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York, NY: Wiley.

Emmerton, J. (1998). Numerosity differences and effects of stimulus density on pigeons's discrimination performance. *Animal Learning and Behavior*, *26*, 243–256.

Fernald, L. D. (1984). *The Hans legacy: a story of science* (1St ed.). Lawrence Erlbaum Assoc Inc.

Galesic, M., Garcia-Rotamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks to low-numeracy people. *Health Psychology*, *28*(210-216).

Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology*, *27*, 787–797.

Gebuis, T., Cohen Kadosh, R., de Haan, E., & Henik, A. (2009). Automatic quantity processing in 5-year olds and adults. *Cognitive Processing*, *10*, 133–142.

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: local diagnostic value. *Arch Neurol Psychiatry*, *44*(2), 398–408.

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

Gigerenzer, G, Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*, 53–96.

Gigerenzer, G, & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*(4), 650–669.

Gigerenzer, G, & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*, 684–704.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592–596.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 513–525.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (Eds.). (1999). *Simple Heuristics That Make Us Smart* (1st ed.). New York: Oxford University Press.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*, 104–122.

Girotto, V., & Gonzales, M. (2008). Children's understanding of posterior probability. *Cognition*, *106*, 325–344.

Goldstein, D., & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review*, *109*, 75–90.

Grady, C. L., & Craik, F. I. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology*, *10*(2), 224–231.

Hagerman, R. J., Jackson, C., Amiri, K., Silverman, A. C., O'Connor, R., & Sobesky, W. (1992). Girls with fragile X syndrome: physical and neurocognitive status and outcome. *Pediatrics*, *89*(3), 395–400.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology*, *93*, 615–626.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*, 1372–1388.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine: Journal of the Association of American Medical Colleges*, *73*(5), 538–540.

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition*, *84*, 343–352.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*, 2261–2262.

Imbo, I., & Vandierendonck, A. (2007). Do multiplication and division strategies rely on executive and phonological working memory resources? *Memory & Cognition*, *35*(7), 1759–1771.

Isaacs, E. B., Edmonds, C. J., Lucas, A., & Gadian, D. G. (2001). Calculation difficulties in children of very low birthweight. *Brain*, *124*(9), 1701 –1707.

Jacobs, C., & Petermann, F. (2005). *Rechenfertigkeiten- und Zahlenverarbeitungs-Diagnostikum für die 2. bis 6. Klasse*. Göttingen: Hogrefe.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, *3*, 281–282.

Kahneman, Daniel, & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454.

Kahneman, Daniel, & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 263–291.

Kaufman, E. L., & Lord, M. W. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 498–525.

Ketelsen, K., & Welsh, M. (2010). Working memory and mental arithmetic: a case for dual central executive resources. *Brain and Cognition*, *74*, 203–209.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358.

Koehler, O. (1950). The ability of birds to count. *Bulletin of Animal Behaviour*, *9*, 41–45.

Kopera-Frye, K., Dehaene, S., & Streissguth, A. P. (1996). Impairments of number processing induced by prenatal alcohol exposure. *Neuropsychologia*, *34*(12), 1187–1196.

Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication - graphical and analog tools. *Annals of the New York Academy of Sciences*, *1128*, 18–28.

Latten, S., Martignon, L., Monti, M., & Multmeier, J. (2011). Die Förderung erster Kompetenzen für den Umgang mit Risiken bereits in der Grundschule: ein Projekt von RIKO-STAT und dem Harding Center. *Stochastik in der Schule*, *31*, 17–24.

Lee, K.-M., & Kang, S.-Y. (2002). Arithmetic operation and working memory: differential supression in dual tasks. *Cognition*, *83*, B63–B68.

Leibniz, G. W. (1677). Toward a universal characteristic. In Leibniz, G W (Ed.), *Selections* (pp. 17–25). New York, NY: Scribner's Sons.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: large-number discrimination in human infants. *Psychological Science*, *14*(5), 396–401.

MacDougall, M. (2009). Dyscalculia, dyslexia, and medical students' needs for learning and using strategies. *Medical Education Online*, *14*, doi: 10.3885/meo.2009.F0000213.

Maliphant, R., Supramaniam, S., & Saraga, E. (1974). Acquiring skill in reading: a review of experimental research. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *15*(3), 175–185.

Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology. General*, *111*(1), 1–22.

Martignon, L., & Krauss, S. (2009). Hands-on activities for fourth graders: a tool box for decision-making and reckoning with risk. *International Electronic Journal of Mathematics Education*, *4*, 227–258.

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior*, *1*(2), 109–121.

Mechner, F., & Guevrekian, L. (1962). Effects of deprivation upon counting and timing in rats. *Journal of Experimental Analysis of Behavior*, *5*, 463–466.

Neumann, J. V., & Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton, USA: Princeton University Press.

Pascal, B., Eliot, T. S., & Trotter, W. F. (2003). *Pensées*. Courier Dover Publications.

Pennington, B. F. (1995). Genetics of learning disabilities. *Journal of Child Neurology;Journal of Child Neurology*, *10*(Suppl 1), S69–S77.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*, 346–354.

Piaget, J. (1971). *Genetic epistemology*. New York: W.W. Norton & Co.

Real, L. A. (1991). Animal choice behavior and the evolution of cognitive architecture. *Science*, *253*, 980–986.

Real, L. A., & Caraco, T. (1986). Risk and foraging in stochastic environments. *Annual Review of Ecology and Systematics*, *17*, 371–390.

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: a review of data and theories. *Subjective probability* (pp. 239–272). New York, NY: Wiley.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: an interim synthesis. *Learning and Individual Differences*, *7*(1), 1–75.

Ruggeri, A., & Katsikopoulos, K. (under review). Make your own kind of cues: when children make more accurate inferences than adults.

Rumbaugh, D. M., Savage-Rumbaugh, S., & Hegel, M. T. (1987). Summation in the chimpanzee (Pan troglodytes). *Journal of Experimental Psychology. Animal Behavior Processes*, *13*(2), 107–115.

Sandrini, M., Rossini, P. M., & Miniussi, C. (2004). The differential involvement of inferior parietal lobule in number comparison: a rTMS study. *Neuropsychologia*, *42*(14), 1902–1909.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, *130*(3), 380–400.

Seitz, K., & Schumann-Hengsteler, R. (2000). Mental multiplication and working memory. *European Journal of Cognitive Psychology*, *12*(4), 552–570.

Shalev, R. S., Manor, O., Kerem, B., Ayali, M., Badichi, N., Friedlander, Y., & Gross-Tsur, V. (2001). Developmental dyscalculia is a familial learning disability. *Journal of Learning Disabilities*, *34*(1), 59–65.

De Smedt, B., Swillen, A., Devriendt, K., Fryns, J. P., Verschaffel, L., & Ghesquière, P. (2007). Mathematical disabilities in children with velo-cardio-facial syndrome. *Neuropsychologia*, *45*(5), 885–895.

Staddon, J. E. R. (1988). Learning as inference. *Evolution and learning*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sternberg, S. (1966). High-speed scanning in human memory. *Science (New York, N.Y.)*, *153*(3736), 652–654.

Sternberg, S. (1969). The discovery of processing stages: extensions of Donder's method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: North-Holland.

Temple, C. M., & Marriott, A. J. (1998). Arithmetical ability and disability in turner's syndrome: A cognitive neuropsychological analysis. *Developmental Neuropsychology*, *14*(1), 47–67.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

Weiss, R. H. (2006). *CFT 20-R: Grundintelligenztestskala 2-Revision*. Göttingen: Hogrefe.

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006). Principles underlying the design of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions: BBF*, *2*, 19.

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*(20).

Woodruff, G., & Premack, D. (1981). Primative mathematical concepts in the chimpanzee: proportionality and numerosity. *Nature*, *293*(5833), 568–570.

Wundt, W. (1880). *Grundzüge der physiologischen Psychologie* ((2) ed.). Leipzig: Engelmann. Retrieved January 7, 2012, from http://vlp.mpiwg-berlin.mpg.de/references?id=lit39751

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, B1–B11.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: a twenty-five year perspective. *Frequency processing and cognition*. New York: Oxford University Press. Retrieved from internal-pdf://Zacks & Hasher (2002) Frequency Processing A twenty-five year perspective-3490361344/Zacks & Hasher (2002) Frequency Processing A twenty-five year perspective.pdf

Zacks, R. T., & Hasher, L. (2007). Human memory. In F. I. Craik & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (pp. 200–230). Mahwah, NJ: Erlbaum.

Zhou, X., Chen, Y., Chen, C., Jiang, T., Zhang, H., & Dong, Q. (2007). Chinese kindergartners' automatic processing of numerical magnitude in Stroop-like tasks. *Memory & Cognition*, *35*(3), 464–470.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: the role of representation in mental computation. *Cognition*, *98*, 287–308.

# Appendix

Appendix A: Study material used in chapter II (original problems)

Appendix B: Study material used in chapter II (revised problems)

Appendix C: Study material used in chapter III

Appendix D: Study material used in chapter IV

## Appendix A: study material used in chapter II (original problems)

As described in the methods section, children were randomly allocated to either the natural frequency or the icon array condition and solved six Bayesian reasoning problems. Children in the icon array condition received the same information as the children in the natural frequency condition, but had an icon array in addition.

Example: „Out of every 12 balls in a box, 8 are soccer balls and 4 are tennis balls. 1 of the 8 soccer balls is white, the other 7 are yellow. 1 of the tennis balls is white, the other 3 are yellow."

1. „If I take a ball out oft he box at random, do you think it is a soccer ball?"
    a. rather yes
    b. rather no

2. Now imagine all the yellow balls in the box. Do you think there are more soccer balls or tennis balls?
    a. more soccer balls
    b. more tennis balls

3. „How many of the yellow balls are soccer balls?"
   ____ out of ____

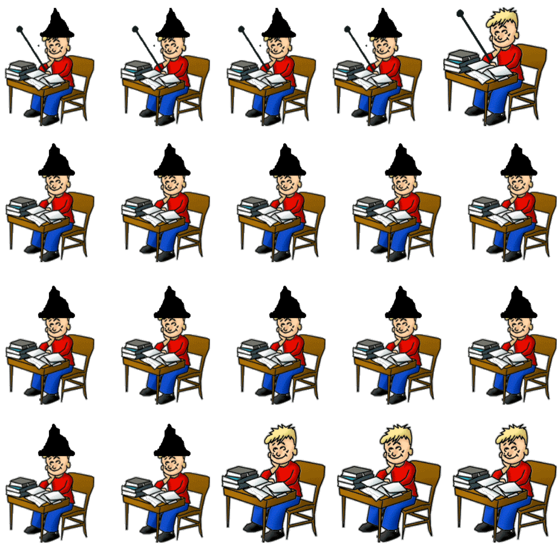
1. “Out of every 30 inhabitants living in a small village, 5 work as firemen, and the other 25 do not work as firemen. 4 of the 5 firemen wear a moustache. Also, 15 of the 25 men who do not work as firemen wear a moustache.”

1. “If I accidently meet someone from that village who is wearing a moustache, do you think he is a fireman?
    a. rather yes
    b. rather no

2. “Now think of all the men wearing moustaches. Are there more firemen, or more men who do not work as firemen?”
    a. more firemen
    b. more men who do not work as firemen

3. “Of all the men with a moustache, how many are firemen?”
    ____ out of ____

2. “Out of every 20 students of Hedgewarth’s school for witchcraft, 5 have a magic stick. Of these 5 students, 4 also wear a magic hat. Of the other 15 students without a magic stick, 12 also wear a magic hat.”



1. “If someone in that school wears a magic hat, do you think he also has a magic stick?”
    a. rather yes
    b. rather no

2. “Imagine a group of students from that school, who are all wearing a magic hat. Are there more students with a magic stick, or more students without a magic stick?”
    a. more students with a magic stick
    b. more students without a magic stick

3. “How many students out of that group who wear a magic hat also have a magic stick?”
    ___ out of ___

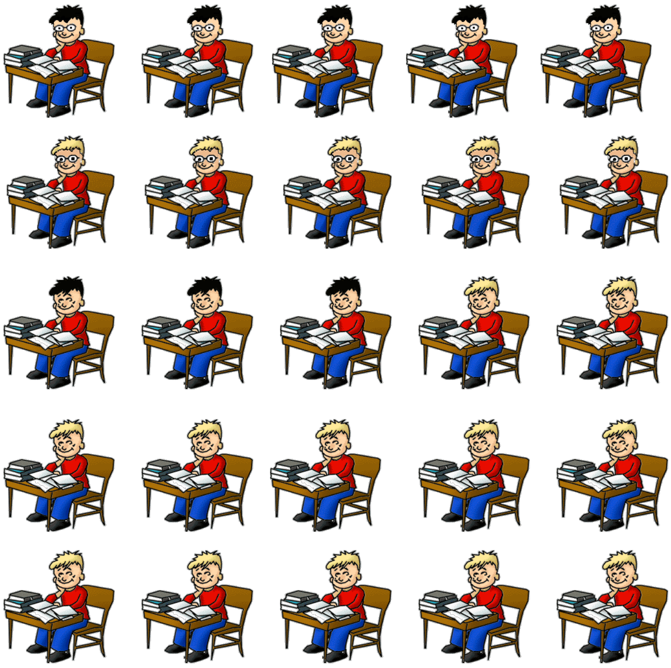
3. “Out of every 10 fairy creatures in a small and far-away fairyland, 2 are princesses, and 8 are mermaids. Of the 2 princesses, 1 wears a crown. Also, 2 of the 8 mermaids wear a crown.”

1. “If I meet someone from that fairyland who is wearing a crown, do you think she is a princess?”
   a. rather yes
   b. rather no

2. “Now think of a group of fairy-creatures who are all wearing crowns. Are there more princesses or mermaids in that group?”
   a. more princesses
   b. more mermaids

3. “Of all the fairy-creatures who are wearing a crown, how many are a princess?”
   ____ out of ____

4. "Of every 25 students in a class, 10 wear glasses, the other 15 do not. Of the 10 students who wear glasses, 5 have dark hair. Of the other 15 students, 3 have dark hair."

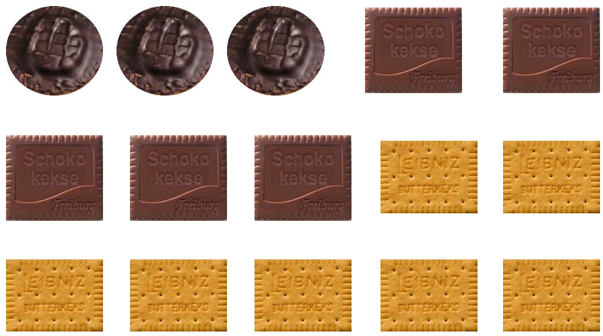


1. "Imagine someone from that class with dark hair. Do you think he has glasses?"
   a. rather yes
   b. rather no

2. "Not imagine a group of students from that class who all have dark hair. Do you think there are more students in that group with glasses, or are there more students without glasses?"
   a. more students with glasses
   b. more students without glasses

3. "How many of the children in the group with dark hair have glasses?"
   ____ out of ____

5. “Out of every 15 cookies in a cookie jar, 3 are round and made of chocolate. The other 12 cookies are square. Of the 12 square cookies, 5 are also made of chocolate.”

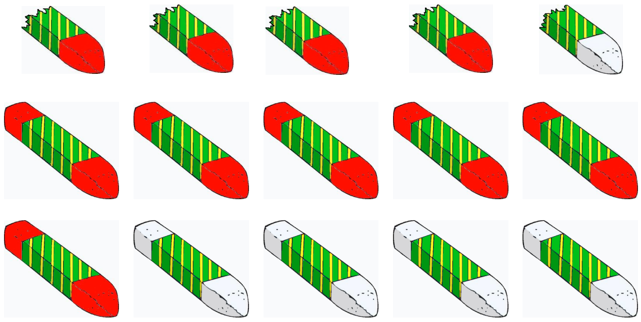



1. “I found a chocolate cookie. Do you think it is round?”
   a. rather yes
   b. rather no

2. “Now picture all the chocolate cookies in the jar. Are more round, or are more square?”
   a. more round
   b. more square

3. “How many of the chocolate cookies are round?”
   ____ out of ____

6. “Out of every 15 chalk sticks in a box, 5 are broken and 10 are whole. Of the 5 broken sticks, 4 are red. Of the 10 whole chalk sticks, 6 are red.”



1. “If I take a red stick out of the box, do you think he will be broken?”
    a. rather yes
    b. rather no

2. “If you think of all the red chalk sticks in the box, are more of them broken, or are more whole?”
    a. more are broken
    b. more are whole

3. “How many of the red chalk sticks are broken?”
    ____ out of ____

## Appendix B: study material used in chapter II (revised problems)

The same six problems described in Appendix A were used for the retest. In problems 1, 5, and 6, questions 1 and 2 were reformulated to incorporate a judgment from one group to another. Changes are indicated in bold font. The graphical information was unchanged and can be found in Appendix A.

1. "Out of every 30 inhabitants living in a small village, 5 work as firemen, and the other 25 do not work as firemen. 4 of the 5 firemen wear a moustache. Also, 15 of the 25 men who do not work as firemen wear a moustache."

    1) "If I accidently meet someone from that village who is wearing a moustache, do you think he is a fireman?
        a. rather yes
        b. rather no

    2) "Now **imagine a group of men from that village** wearing moustaches. Are there more firemen, or more men who do not work as firemen **in that group**?"
        a. more firemen
        b. more men who do not work as firemen

    3) "Of all the men with a moustache **in the group**, how many are firemen?"
    ____ out of ____

5. "Out of every 15 cookies in a cookie jar, 3 are round and made of chocolate. The other 12 cookies are square. Of the 12 square cookies, 5 are also made of chocolate."

    1) "I found a chocolate cookie. Do you think it is round?"
        rather yes
        rather no

    2) "Now **imagine you take out some chocolate cookies from the jar**. Are more **of them** round, or are more square?"

more round

more square

3) "How many of the chocolate cookies **that you took out of the jar** are round?"

____ out of ____

6. "Out of every 15 chalk sticks in a box, 5 are broken and 10 are whole. Of the 5 broken sticks, 4 are red. Of the 10 whole chalk sticks, 6 are red."

1) "If I take a red stick out of the box, do you think he will be broken?"
    a. rather yes
    b. rather no

2) "**Imagine you take some red chalk sticks out of the box**, are more of them broken, or are more whole?"
    a. more are broken
    b. more are whole

3) "How many of the red chalk sticks **that you have taken from the box** are broken?"

____ out of ____

## Appendix C: study material used in chapter III

### Icon array condtion

Example: „A box contains several balls. Here you see what they could look like.“


1) Now imagine you take some yellow balls out of the box. Do you think there are more soccer balls or tennis balls?
   a. more soccer balls
   b. more tennis balls

2) "Now take a green pencil and circle every yellow ball. then take a red pencil and cross out every yellow soccer ball."

3) „How many of the yellow balls that you took out of the box are soccer balls?“
   ____ out of ____

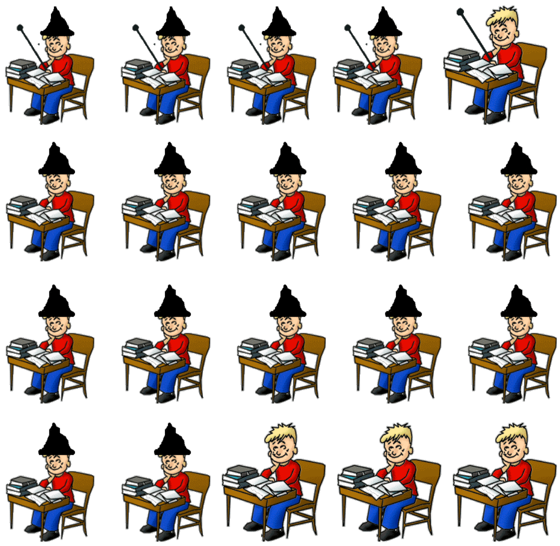
1. "In a small village, some inhabitants work as firemen. Some of the firemen wear moustaches. Also, some of the men who do not work as firemen wear moustaches. Here you see some of the people in the village."

1) "Now imagine you meet a group of people from the village who all wear moustaches. Are there more firemen, or more men who do not work as firemen in that group?"
   a. more firemen
   b. more men who do not work as firemen

2) "Now take a green pencil and circle every man with a moustache. Then take a red pencil and cross out every fireman who wears a moustache."

3) "Of the men with a moustache in the group, how many are firemen?"
   ____ out of ____

2. "Of the students of Hedgewarth's school for witchcraft, some have a magic stick. Of these, some also wear a magic hat. Also some of the students who do not have a magic stick wear a magic hat. Here you see some of the children from that school."



1) "Imagine a group of students from that school, who are all wearing a magic hat. Are there more students with a magic stick, or more students without a magic stick?"
    a. more students with a magic stick
    b. more students without a magic stick

2) "Now take a green pencil and circle student with a magic hat. Then take a red pencil and cross out every student who wear a magic hat and have a magic stick."

3) "How many students out of that group who wear a magic hat also have a magic stick?"
    ____ out of ____

3. "In a far-away fairyland, some inhabitants are princesses, the other are mermaids. Some of the fairy-creatures wear a crown. Here you can see some of them."

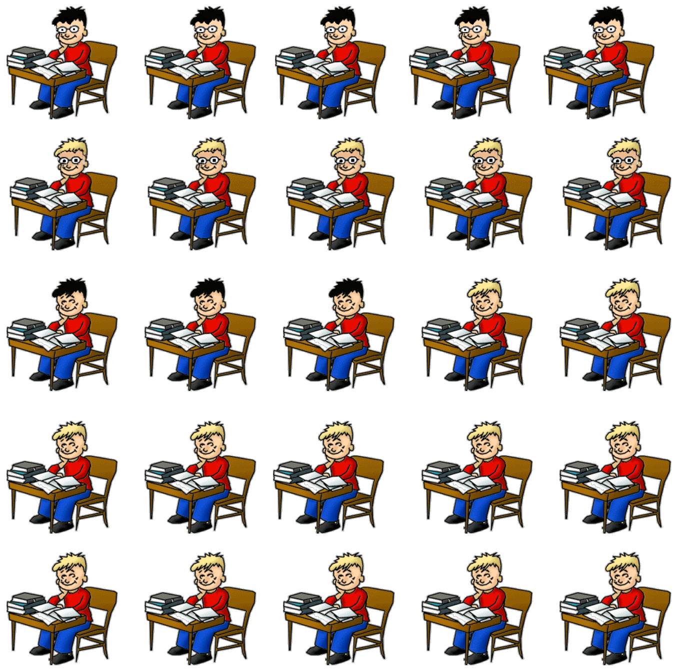
1) "Now think of a group of fairy-creatures who are all wearing crowns. Are there more princesses or mermaids in that group?"
   a. more princesses
   b. more mermaids

2) "Now take a green pencil and circle every fairy-creature who is wearing a crown. Then take a red pencil and cross out every princess that has a crown."

3) "Of all the fairy-creatures who are wearing a crown, how many are a princess?"

   ____ out of ____

4. “Some students in a class have glasses, the others do not. Some of the students who wear glasses have dark hair. Also, some of the children who do not wear glasses have dark hair. Here you can see some of the students.”

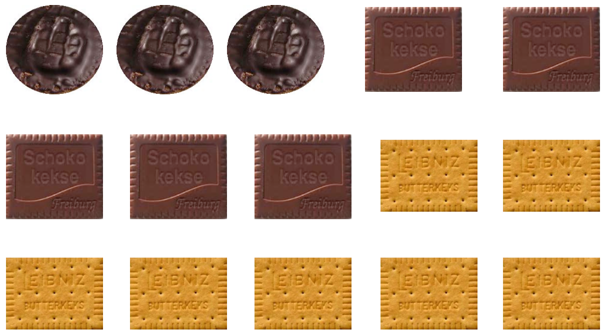


1) “Not imagine a group of students from that class who all have dark hair. Do you think there are more students in that group with glasses, or are there more students without glasses?”
    a. more students with glasses
    b. more students without glasses

2) “Now take a green pencil and circle every student who has dark hair. Then take a red pencil and cross out student that has dark hair and wears glasses.”

3) “How many of the children in the group with dark hair have glasses?”
    ____ out of ____

5. "A cookie jar contains some cookies. Some of the cookies are round and made of chocolate. The others are square. Of the square cookies, some are also made of chocolate. Here are a few cookies from the jar."

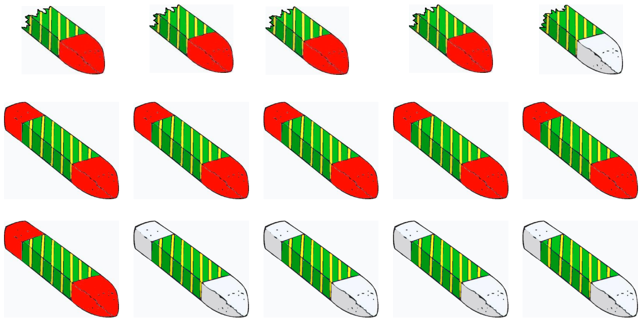


1) "Imagine you take some chocolate cookies from the jar. Are more round, or are more square cookies?"
    a. more round
    b. more square

2) "Now take a green pencil and circle every chocolate cookie. Then take a red pencil and cross out student every chocolate cookie that is round."

3) "How many of the chocolate cookies are round?"
    ___ out of ___

6. “A box contains sticks of chalk. Some sticks are broken, the others are whole. Some of the broken sticks are red. Also, some of the whole sticks are red. Here you see some of the sticks from the box.”



1) “If you take out some red chalk sticks from the box, are more of them broken, or are more whole?”
    a. more are broken
    b. more are whole

2) “Now take a green pencil and circle every red chalk stick. Then take a red pencil and cross out student every red chalk stick that is broken.”

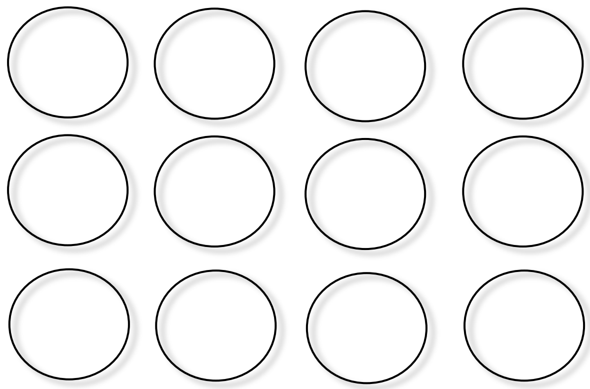
3) “How many of the red chalk sticks are broken?”
   ____ out of ____

**Natural frequency condition**

Example: „Out of every 12 balls in a box, there are

- 1 white soccer ball
- 7 yellow soccer balls
- 1 white tennis balls
- 3 yellow tennis balls

Here you see 12 balls from that box. But which are white and which are yellow? And which are soccer and which are tennis balls?“



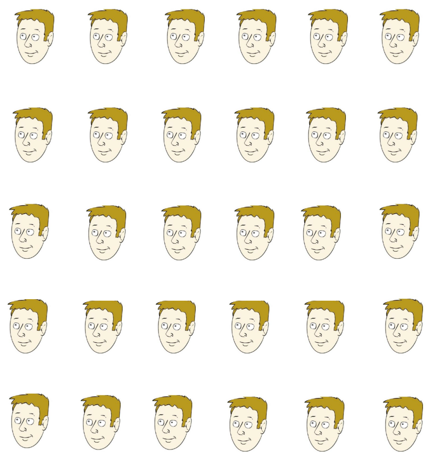
1) “Now imagine you take some yellow balls out of the box. Do you think there are more soccer balls or tennis balls?
   a. more soccer balls
   b. more tennis balls

2) “Now look at the balls above and take a green pencil and circle every ball that would be yellow. Then take a red pencil and cross out every ball that would be a yellow soccer ball. Everything you need is in the text above.”

3) „How many of the yellow balls that you took out of the box are soccer balls?“
   ____ out of ____

1. "Out of each 30 men living in a village, there are
   - 4 firemen who wear a moustache
   - 1 fireman who does not wear a moustache
   - 15 men who do not work as firemen and wear a moustache
   - 10 men who do not work as firemen and do not have a moustache

   Here you see 30 people from that village. But who wears a moustache, and who works as fireman?"



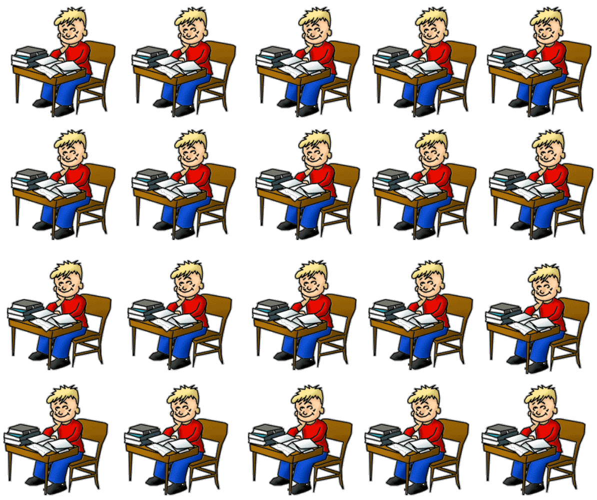
1) "Now imagine you meet a group of people from the village who all wear moustaches. Are there more firemen, or more men who do not work as firemen in that group?"
   a. more firemen
   b. more men who do not work as firemen

2) "Now take a green pencil and circle every man who would have a moustache. Then take a red pencil and cross out every fireman who would wear a moustache."

3) "Of the men with a moustache in the group, how many are firemen?"
   ____ out of ____

2. "Out of every 20 students in Hedgeworth's school for witchcraft,
    - 5 have a magic stick and wear a magic hat
    - 1 has a magic stick, but no magic hat
    - 12 do not have a magic stick, but wear a magic hat
    - 3 do not have a magic stick, and also do not wear a magic hat

    Here you see 20 of the students. But who has a magic stick, and who wears a magic hat?"



1) "Imagine a group of students from that school, who are all wearing a magic hat. Are there more students with a magic stick, or more students without a magic stick?"
    a. more students with a magic stick
    b. more students without a magic stick

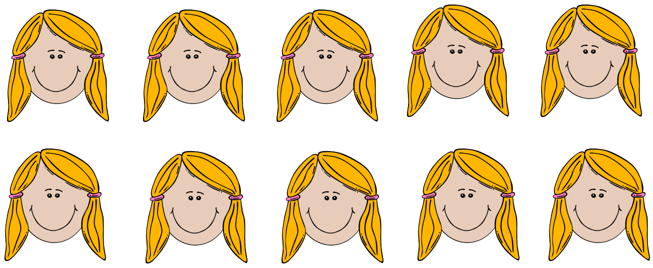
2) "Now take a green pencil and circle student with a magic hat. Then take a red pencil and cross out every student who wear a magic hat and have a magic stick."

3) "How many students out of that group who wear a magic hat also have a magic stick?"

    ____ out of ____

3. “Out of every 10 fairy-creatures in a far-away fairyland, there are
   - 1 princess who is wearing a crown
   - 1 princess who is not wearing a crown
   - 2 mermaids who are wearing a crown
   - 6 mermaids who are not wearing a crown

   Here you see 10 of the fairy-creatures. But who is a princess, and who is wearing a crown?”



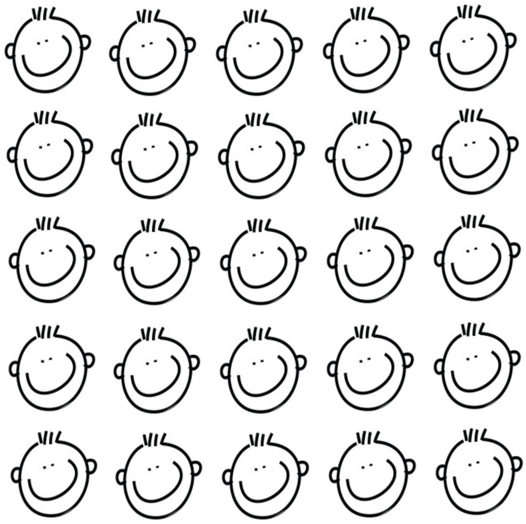
1) “Now think of a group of fairy-creatures who are all wearing crowns. Are there more princesses or mermaids in that group?”
   a. more princesses
   b. more mermaids

2) “Now take a green pencil and circle every fairy-creature who is wearing a crown. Then take a red pencil and cross out every princess that has a crown.”

3) “Of all the fairy-creatures who are wearing a crown, how many are a princess?”
   ____ out of ____

4. "Out of every 25 students of one class,
   - 5 students wear glasses and have dark hair
   - 5 students wear glasses and are blond
   - 3 students do not wear glasses and have dark hair
   - 12 students do not wear glasses and are blond.

   Here you see 25 students from that class. Do you know who wears glasses, and who has dark hair?"



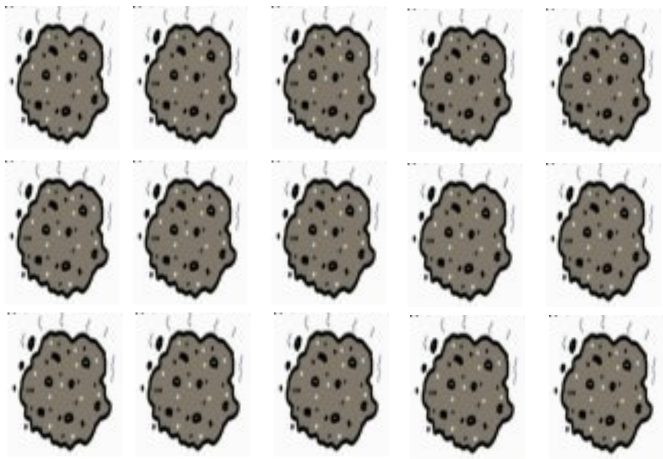
1) "Not imagine a group of students from that class who all have dark hair. Do you think there are more students in that group with glasses, or are there more students without glasses?"
   a. more students with glasses
   b. more students without glasses

2) "Now take a green pencil and circle every student who has dark hair. Then take a red pencil and cross out student that has dark hair and wears glasses."

3) "How many of the children in the group with dark hair have glasses?"
   ____ out of ____

5. “Out of every 15 cookies in a cookie jar,
    - 3 are round and made of chocolate
    - 5 are square and made of chocolate
    - 7 are square and whole-grain

Here you see 15 cookies. But which are made of chocolate, and which are round?”



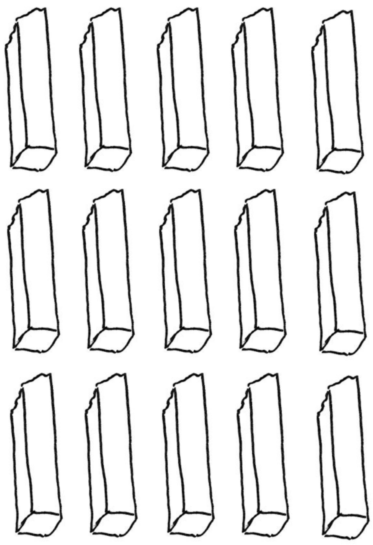
1) “Imagine you take some chocolate cookies from the jar. Are more round, or are more square cookies?”
    a. more round
    b. more square

2) “Now take a green pencil and circle every chocolate cookie. Then take a red pencil and cross out student every chocolate cookie that is round.”

3) “How many of the chocolate cookies are round?”
    ____ out of ____

6. "Out of every sticks of chalk in a box, there are
   - 4 which are broken and red
   - 1 which is broken and white
   - 6 which are not broken and red
   - 4 which are not broken and white

   Here you see a picture of 15 chalk sticks. But which ones are red, and which ones are broken?"



1) "If you take out some red chalk sticks from the box, are more of them broken, or are more whole?"
   a. more are broken
   b. more are whole

2) "Now look at the picture above. Take a green pencil and circle every red chalk stick in the box. Then take a red pencil and cross out student every red chalk stick that would be broken."

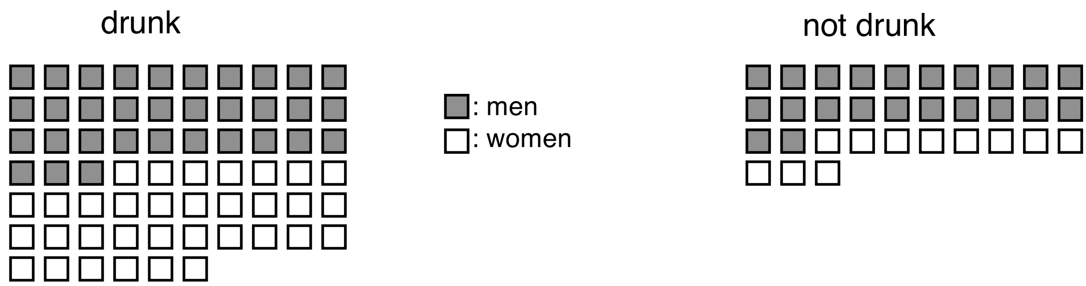
3) "How many of the red chalk sticks that you took out of the box before would be broken?"
   ____ out of ____

## Appendix D: study material used in chapter IV

### Problem 1

On one weekend, the police checks cars for drunk drivers. Some of the drivers were drunk. Here are some of the drivers the police checked:

Icon array condition:




Natural frequency condition:

- 66 of the drivers were drunk
- 33 of the drunk drivers were men
- 33 of the drivers were not drunk
- 22 of the drivers who were not drunk were men

Conditional probability condition:

- 66% of the drivers were drunk
- 50% of the drunk drivers were men
- 33% of the drivers were not drunk
- 66% of the drivers who were not drunk were men

In a different group of men, were there

❑ more drunk men, or

❑ more men who were not drunk?

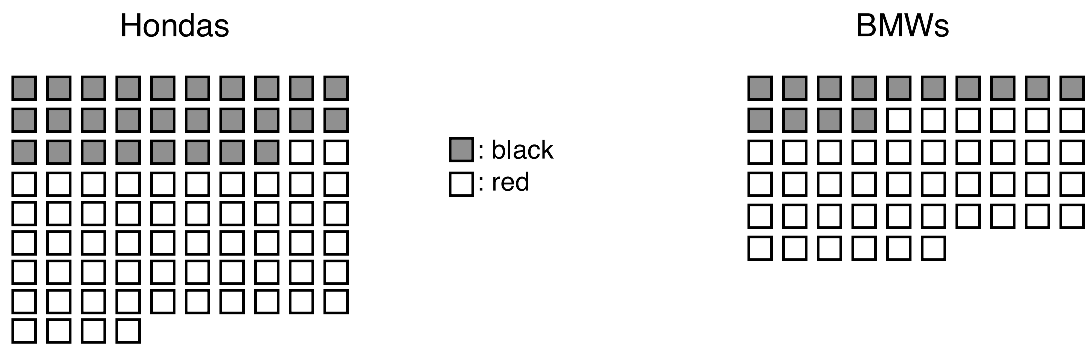
How many men in that group were drunk?

____ out of ____

## Problem 2

Christian is counting motorbikes on the highway. Some motorbikes were black, the others were red. Also, the motorbikes were either BMWs or Hondas. Here are some of the bikes Christian counted:

Icon array condition:




Natural frequency condition:

- 84 of the motorbikes were Hondas
- 28 of the Hondas were black
- 56 of the motorbikes were BMWs
- 14 of the BMWs were black

Conditional probability condition:

- 60% of the motorbikes were Hondas
- 33% of the Hondas were black
- 40% of the motorbikes were BMWs
- 25% of the BMWs were black

In another group of black motorbikes, were there

❑ more Hondas, or

❑ more BMWs?

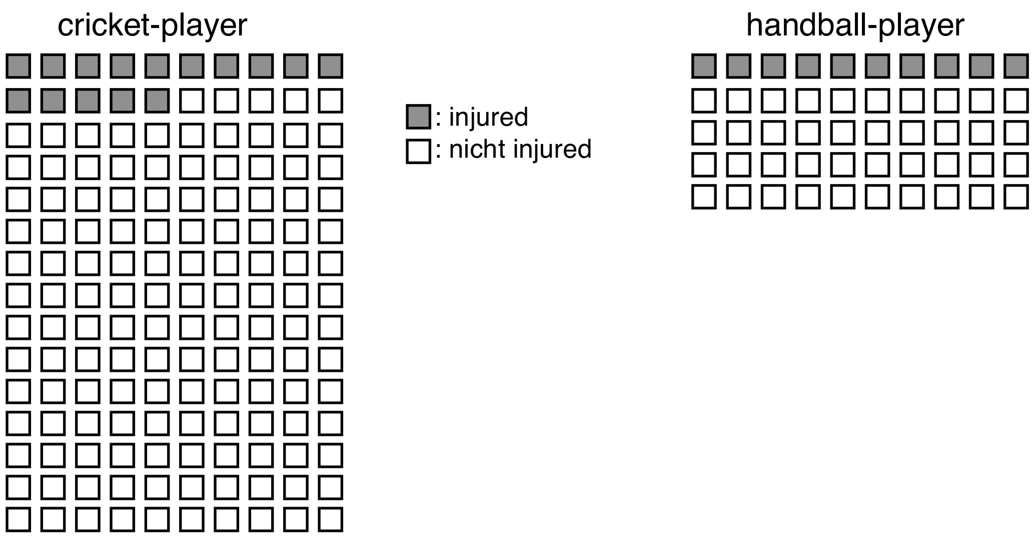
How many of the black motorbikes in that group were Hondas?

____ out of ____

**Problem 3**

In the city sports club of Smallville, people play either handball or cricket. Some of the people sustained sports injuries in the last year. Here are some of the people who hurt themselves playing sports:

Icon array condition:




Natural frequency condition:

- 150 of the people play cricket
- 15 of the people who play cricket got injured last year
- 50 of the people play handball
- 10 of the people who play handball got injured last year

Conditional probability condition:

- 80% of the people play cricket
- 20% of the people who play cricket got injured last year
- 20% of the people play handball
- 40% of the people who play handball got injured last year

Imagine a group of people who hurt themselves playing sports last year. Did more people get hurt playing

❒ cricket, or

❒ handball?

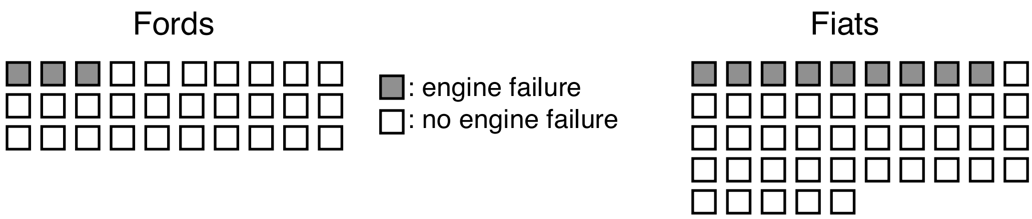
In that group of people who had a sports injury last year, how many played handball?

____ out of ____

**Problem 4**

In Huntsville, the people either drive a Ford or a Fiat. Some of the cars had to be repaired for engine failure last year. Here are some of the cars in Huntsville:

Icon array condition:




Natural frequency condition:

- 30 of the cars were Fords
- 3 of the Fords had engine failures
- 45 of the cars were Fiats
- 9 of the Fiats had engine failures

Conditional probability condition:

- 40% of the cars were Fords
- 10% of the Fords had engine failures
- 60% of the cars were Fiats
- 20% of the Fiats had engine failures

Imagine some cars with engine failures. Would there be

❒ more Fords, or

❒ more Fiats?

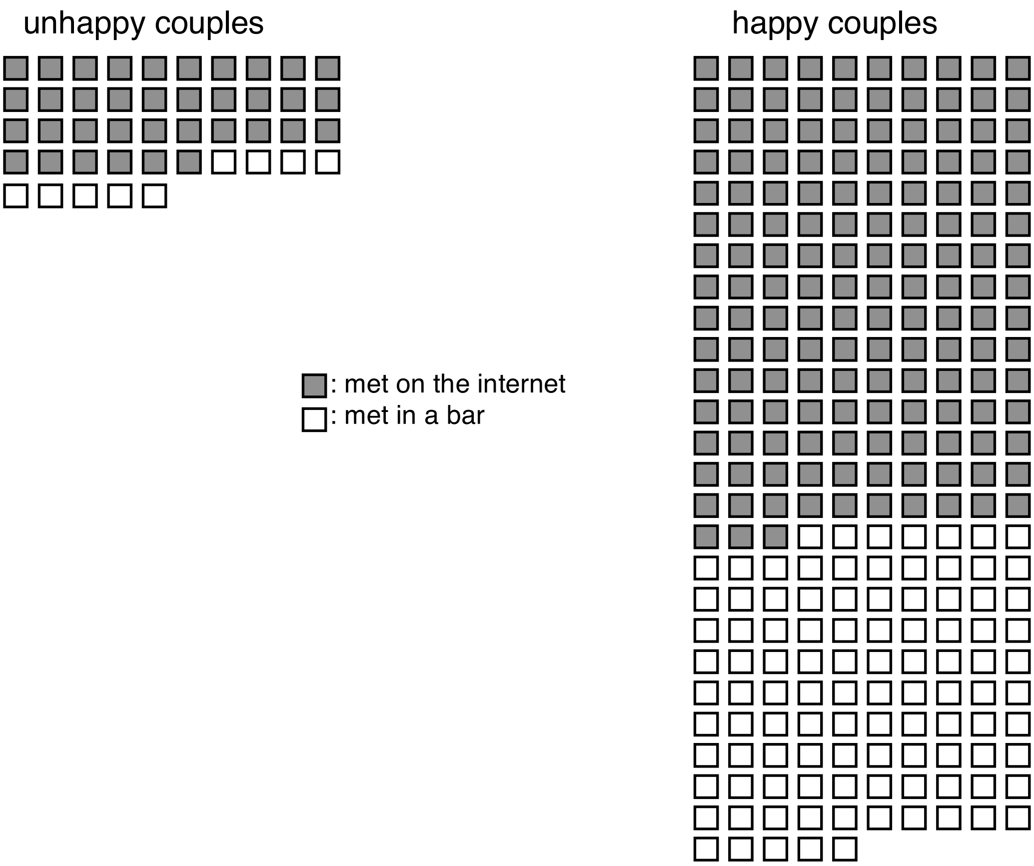
How many of the cars with engine failures were Fords?

____ out of ____

**Problem 5**

In Santa Sofia, couples either met on the internet or in a bar. Some of the couples are still happy with their relationship, the others are not. Here are some of the couples

Icon array condition:




Natural frequency condition:

- 255 of the couples are happy with their relationship
- 153 of the happy couples met on the internet
- 45 of the couples are not happy with their relationship
- 36 of the unhappy couples met on the internet

Conditional probability condition:

- 80% of the couples are happy with their relationship
- 60% of the happy couples met on the internet
- 20% of the couples are not happy with their relationship
- 80% of the unhappy couples met on the internet

In a group of couples who met on the internet, are

❒ more happy, or

❒ more unhappy?

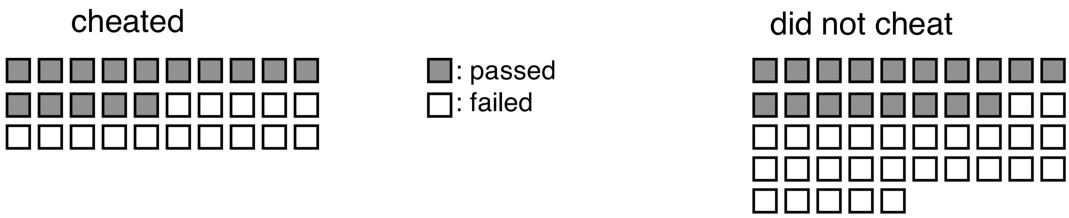
How many couples of those in the group who met on the internet are happy?

____ out of ____

**Problem 6**

Students have to take a final exam after a university course. Some of the students pass, the others fail the exam. Also, some students cheated on the exam. Here you see a group of students:

Icon array condition:




Natural frequency condition:

- 30 of the students cheated on the exam
- 15 of the students who cheated passed the exam
- 45 of the students did not cheat on the exam
- 18 of the students who did not cheat passed the exam

Conditional probability condition:

- 20% of the students cheated on the exam
- 40% of the students who cheated passed the exam
- 80% of the students did not cheat on the exam
- 20% of the students who did not cheat passed the exam

Ina group of students who all passed the exam, did

❒ more of the cheat, or

❒ did more of them not cheat?

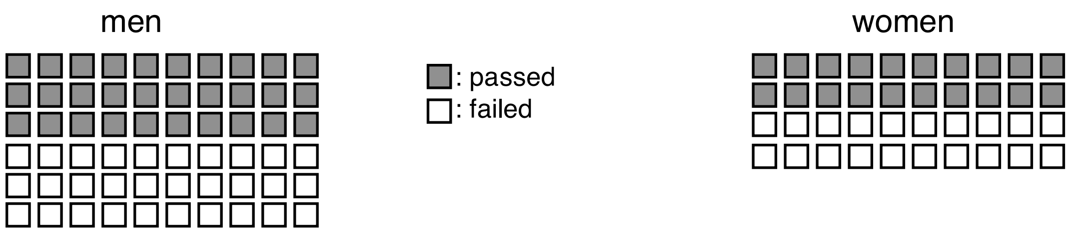
How many students in the group who all passed their exam cheated?

____ out of ____

**Problem 7**

Some students in a driver's education class passed their exams last year, the others failed. Here is a group of students from that class:

Icon array condition:




Natural frequency condition:

- 60 of the students were male
- 30 of the male students passed the exam
- 40 of the students were female
- 20 of the female students passed the exam

Conditional probability condition:

- 60% of the students were male
- 50% of the male students passed the exam
- 40% of the students were female
- 50% of the female students passed the exam

In a group of students who passed the driver's education exam last year, are there

❒ more women, or

❒ more men?

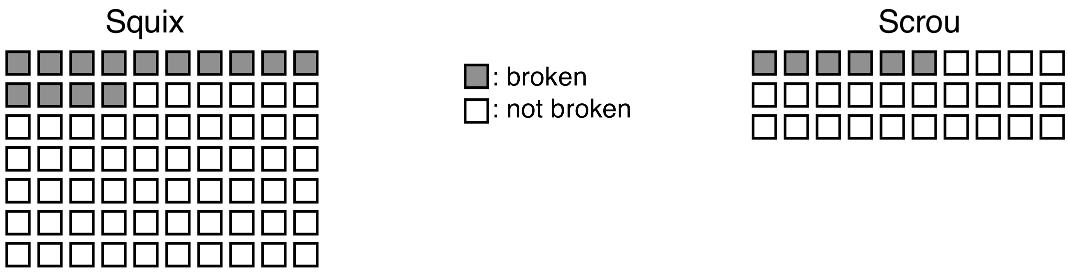
How many students in a group who passed their exam last year were male?

____ out of ____

## Problem 8

A hardware store sells screws of two brands, “Squix” and “Scrou”. Some of the screws are broken. Here you see a sample of the screws:

Icon array condition:




Natural frequency condition:

- 70 of the screws are made from “Squix“
- 14 of the “Squix” screws are broken
- 30 of the screws are made from “Scrou”
- 6 of the “Scrou” screws are broken

Conditional probability condition:

- 70% of the screws are made from “Squix“
- 20% of the “Squix” screws are broken
- 30% of the screws are made from “Scrou”
- 20% of the “Scrou” screws are broken

In a sample of screws which are all broken, are

❒ more from “Squix”, or

❒ more from “Scrou”?

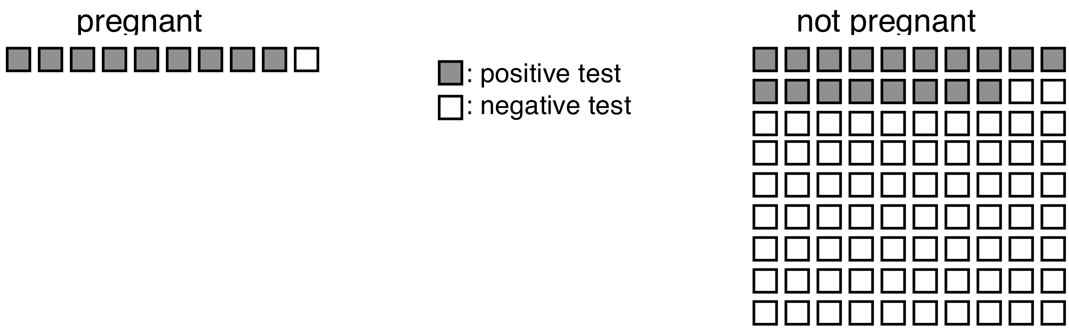
How many screws in that sample are made from “Squix”?

____ out of ____

**Problem 9**

Some of the women who take a Rite-Aid pregnancy test and are pregnant have a positive test result. Also, some of the women who are not pregnant test positive on the test. Here is a group of women who took the test:

Icon array condition:




Natural frequency condition:

- 10 of the women are pregnant
- 9 of the pregnant women test positive
- 90 of the women are not pregnant
- 18 of the women who are not pregnant test positive

Conditional probability condition:

- 10% of the women are pregnant
- 90% of the pregnant women test positive
- 90% of the women are not pregnant
- 20% of the women who are not pregnant test positive

In a group of women who all received a positive test result, are

❐ more women pregnant, or

❐ are more women not pregnant?

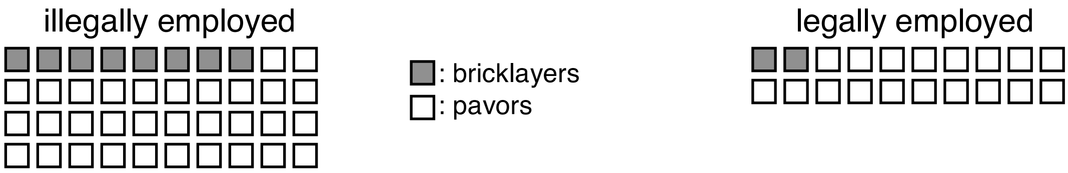
How many of the women in the group who received a positive test result are pregnant?

____ out of ____

**Problem 10**

On a large construction site, people either work as bricklayers or pavors. The police conducts a control to find workers who are not legally employed. Here are some of the workers who were controlled:

Icon array condition:




Natural frequency condition:

- 40 of the workers were illegally employed
- 8 of the illegal workers were bricklayers
- 20 of the workers were legally employed
- 2 of the legal workers were bricklayers

Conditional probability condition:

- 60% of the workers were illegally employed
- 20% of the illegal workers were bricklayers
- 40% of the workers were legally employed
- 10% of the legal workers were bricklayers

In a group of bricklayers, were there

❑ more who were legally employed, or

❑ more who were illegally employed?

In the group of bricklayers, how many were illegally employed?

____ out of ____

# Summary

Bayesian reasoning describes the process of assessing the likelihood of a hypothesis based on the prior probability of that hypothesis and new information which may strengthen or weaken the hypothesis, called the posterior probability. In medical settings, Bayesian reasoning is involved when making inferences about a disease such as breast cancer given a positive test result, e.g., from a routine mammography screening. Previous prominent studies on Bayesian reasoning postulated that humans were cognitively biased, relied on misleading heuristics when making inferences, and finally were "not Bayesian at all". More recently, however, research has focused on the *external* representation of the problem, instead of searching for explanations inside of the mind only. By reformulating the information from conditional probabilities into natural frequencies, rates of correct estimates of posterior probabilities increased drastically. Two explanations for this effect have been proposed: 1) natural frequencies facilitate the computation of Bayesian posterior probabilities because they demand fewer arithmetic operations, and 2) natural frequencies mimic the way information is sampled in natural environments (i.e. they follow an ecological design), and therefore conform to the way humans perform Bayesian reasoning in their every-day lives.

In this dissertation, I test both claims more thoroughly, focusing on quantity representations and the implications different representations have for the arithmetic complexity of Bayesian reasoning operations and working-memory involvement. In the real world, what we observe are series of real, discrete, and countable events. While the ability to understand number symbols like Arabic

numerals is specific for humans and requires cultural education, humans and other animals possess a sense for discrete and countable quantities, termed *analogue* quantities, Those quantities can be directly perceived by the senses and are not transformed in a verbal or visual code.

Chapter II investigates if children in second and fourth grades of elementary school can solve Bayesian reasoning problems better, if the information is presented as analogue quantities instead of natural frequencies. Also, in addition to precise estimates of posterior probabilities, children were asked for categorical inferences about the hypothesis, which assess Bayesian reasoning performance at a coarser level. Results suggest that presenting information as discrete, countable objects increases the number of correct solutions, and makes more children choose a Bayesian strategy systematically to answer the problems. The level of correct answers for categorical inferences was higher than that for precise estimates of posterior probabilities. Younger children were more accurate in judging the frequency of events than the probability of a single event, while no difference was found for older children. Apparently, even second-graders can systematically engage in Bayesian reasoning if the information is presented in a way that children perceive it in their natural environments. A follow-up study was conducted because some questions in the original study did not demand inferences from one set of events to another. The follow-up study followed the same design as the original, except that the task and question formulation was changed to involve making inferences. The same differences as reported in the first study were observed, so that the results reported earlier were not due to flawed question formats.

Chapter III extends the results observed in chapter II for regular children to those with clinically impaired arithmetic competencies, i.e. developmental dyscalculia. Developmental dyscalculia is an isolated deficit in processing numerical material and solving arithmetical problems. Some children develop dyscalculia because they lack the most basic sense of quantity due to genetic preconditions or adverse intrauterine events. Those children show a severe deficit in processing single, countable events, as well as number symbols. Other children, however, develop dyscalculia because they have linguistic processing deficits which prevent them from developing a reliable connection between discrete events and symbolic number representations. While those children have a deficit in encoding the meaning of number symbols, their basic quantity representations remain intact. Children with a confirmed diagnosis of dyscalculia solved Bayesian reasoning problems with either symbolic quantity representations (natural frequencies), or analogue quantities. Children could solve more problems with analogue quantity information than with natural frequencies. Subsequent analysis showed that the beneficial effect of analogue quantities was influenced by the intactness of basic quantity representations in children. While analogue quantity information was generally helpful for children with dyscalculia, in line with clinical models of dyscalculia, children with impaired basic quantity representations can extract fewer meaningful information from analogue quantities, compared to children who only experience problems in processing symbolic quantities.

Finally, chapter IV investigated if the differences found in the number of correct solutions on Bayesian reasoning tasks with different representations can also be detected in response times, and if the different arithmetic operations involved recruit different working-memory resources. As the number of arithmetic steps increases, not only does the probability of errors increase, but also the time it takes to determine the correct answer. In the first part of this chapter, response times were assessed for Bayesian reasoning tasks with different representations. Mirroring the number of correct solutions, response times increased with increasing processing demands. While analogue quantities led to fast categorical judgments, the time needed to answer questions about posterior probabilities depended on the number of events which had to be counted. Natural frequencies showed fast response times for both, categorical judgments and judgments for posterior probabilities. With conditional probability information, response times were high for judgments, regardless of the precision.

Not only the number of computational steps is different between the task representations, but also their quality: while natural frequencies involve performing one addition, conditional probabilities rely on multiplications to determine the answer. Analogue quantities require the single events to be counted to calculate the posterior probability. Addition involves the manipulation of quantities on a mental number line, and recruits visuo-spatial working-memory resources. On the other hand, multiplications and counting rely on the verbal-auditory working-memory system. The second part of chapter IV reports the results of a dual-task study in which participants had to perform Bayesian reasoning tasks with different representations at the same time as reactions tasks either blocking resources of the verbal-auditory or visuo-spatial working-memory system. As predicted, Bayesian reasoning performance with natural frequencies was affected by a visuo-spatial, but not a verbal-auditory secondary task, while the opposite pattern emerged for icon arrays and conditional probabilities. Also, while making categorical inferences was susceptible to secondary task interference for natural frequencies and conditional probabilities, neither the rate of correct solutions nor response times declined under dual-task conditions for analogue quantities. Possibly extracting meaning from analogue quantities occurs without intention and relies less on working-memory resources.

Overall, the research presented in this dissertation supports both claims – that presenting information the way it is encountered in natural environments increases reasoning performance, and that differences in computational load determine the rate of correct solutions for Bayesian reasoning problems with different solutions. But besides confirming assertions which have been made over a decade ago, this work hopefully contributed a number of new elements to theories of Bayesian reasoning. First, I could show that even children in second grade can systematically integrate information according to Bayes rule, when they are given quantity representations that they can apprehend. Second, even children with severe impairments in performing arithmetic operations can reason the Bayesian way when the information does not involve number symbols. Also, this effect is stronger for children with intact representations of analogue quantities. And finally, this dissertation for the first time reported response time data for Bayesian reasoning problems. Given the long history of response time analysis as a tool for assessing mental workload and differentiating cognitive processes, they should compliment rates of correct responses in future studies interested in the mental mechanisms that underlie Bayesian reasoning. Likewise, for the first time a dual-task approach was used to distinguish working-memory demands of the different representation formats.

Patterns of interference were generally in line with previous research in numerical cognition.

To conclude, this dissertation extended the previous assertions of computational facilitation and ecological design, and points out links to theories of the development and impairment of quantity representation, and research connecting cognitive arithmetic to elementary processes of working-memory. Hopefully, this research will prove useful in deepening these connections.

# Zusammenfassung (German Summary)

Bayesianisches Inferenzen sind Schlussfolgerungen über die Wahrscheinlichkeit einer Hypothese anhand der *a-priori*-Wahrscheinlichkeit der Hypothese, und neuen Informationen, die für oder gegen die Hypothese sprechen können. Die resultierende Wahrscheinlichkeit nach Berücksichtigung neuer Informationen wird auch *a-posteriori*-Wahrscheinlichkeit genannt.

Bayesianische Inferenzen finden beispielsweise in der Medizin Anwendung, wenn aufgrund eines positiven Testergebnis auf das Vorhandensein einer Krankheit geschlossen wird, z.B. bei einer Routine-Mammographie im Rahmen eines Screeningprogramms für Brustkrebs. Erste Untersuchungen der kognitiven Psychologie kam zu der Schlussfolgerung, dass Menschen sich bei Inferenzen auf irreführende Heuristiken stützen, die zu einer Verzerrung der Urteile führen. Neuere Studien heben jedoch die Rolle der Repräsentation der Informationen hervor, statt pauschal ein kognitives Defizit zu attestieren. Durch eine Umformulierung der Informationen von bedingten Wahrscheinlichkeiten zu natürlichen Häufigkeiten erhöhte sich die Rate korrekter Bayesianischer Inferenzen drastisch. Zwei Erklärungen für diesen Effekt wurden vorgeschlagen: 1) natürliche Häufigkeiten erleichtern die Berechnung der *a-posteriori*-Wahrscheinlichkeiten, weil sie weniger Rechenoperationen benötigen, und 2) natürliche Häufigkeiten imitieren, wie Menschen Informationen in natürlichen Umgebungen aufnehmen (d.h. natürliche Häufigkeiten folgen einem ökologischen Design), und damit der Art und Weise, wie Menschen Inferenzen im alltäglichen Leben vornehmen.

In dieser Dissertation untersuche ich beide Erklärungsansätze genauer. Dabei liegt der Fokus zum Einen auf der normalen bzw. verzögerten Entwicklung von Mengenrepäsentation und darauf, wie früh entwickelte, analoge Mengenrepräsentationen Bayesianische Inferenzen im Vergleich zu symbolischen Mengenrepräsentationen verbessern. Die Fähigkeit, Symbole wie arabische Ziffern zu verstehen ist spezifisch für den Menschen und erfordert kulturelle Bildung. Andererseits haben Menschen und andere Tiere einen Sinn für diskrete und zählbaren Mengen. Diese sogenannten analogen Mengen können direkt mit den Sinnen wahrgenommen werden und sind nicht in einen verbalen oder visuellen Code transformiert.

Zum Anderen liegt der Fokus dieser Arbeit auf den Auswirkungen verschiedener Repräsentationen auf die arithmetische Komplexität der Berechnung der Bayesianischen *a-posteriori*-Wahrscheinlichkeit. Die Komplexität der Berechnung sollte sich nicht nur in der Anzahl korrekter Inferenzen widerspiegeln, sondern auch Auswirkungen auf die Reaktionszeit und die Anforderungen an Arbeitsgedächtnisprozesse haben.

Kapitel II untersucht, ob Kinder in der zweiten und vierten Klasse der Grundschule Bayesianische Inferenzaufgaben besser lösen, wenn die Informationen als analoge Mengen anstelle von natürlichen Häufigkeiten präsentiert werden. Neben der genauen Schätzungen der *a-posteriori*-Wahrscheinlichkeiten sollten die Kinder kategoriale Inferenzen bezüglich der Hypothese vornehmen. Diese kategorialen Inferenzen erfordern lediglich ein Urteil, ob die Hypothese eher wahrscheinlich oder eher unwahrscheinlich ist, und erfassen damit Bayesianische Denkprozesse auf einem gröberen Niveau. Die Ergebnisse zeigen, dass die Darstellung von Informationen als analoge Mengen zu mehr korrekten Inferenzen führt und mehr Kinder systematisch eine Bayesianische Strategie anwenden, um die Aufgaben zu beantworten. Das Niveau der richtigen Antworten für kategoriale Inferenzen war höher als das für präzise Schätzungen der *a-posteriori*-Wahrscheinlichkeiten. Jüngere Kinder konnten die Häufigkeit von Ereignissen besser einschätzen als die Wahrscheinlichkeit einzelner Ereignisse, während ältere Kinder beide Arten von Ereignissen gleich gut beurteilen konnten. Offenbar können auch Zweitklässler systematisch eine Bayesianische Inferenzstrategie anwenden, wenn die Informationen so dargeboten werden, wie sie Kinder in ihrer natürlichen Umgebung wahrnehmen. Eine Nachfolgeuntersuchung wurde durchgeführt, weil einige Fragen in der ursprünglichen Studie keine Inferenzen von der einen auf eine andere Gruppe erforderten. Die Folgeuntersuchung hatte den gleichen Aufbau wie das Original, außer, dass die Formulierung der Aufgabe und Fragen geändert wurden, so dass alle Aufgaben nun Inferenzen erforderten. In dieser Folgeuntersuchung ergaben sich die gleichen Effekte wie in der ursprünglichen Studie berichtet, so dass die Aufgabenformulierung nicht für die berichteten Ergebnisse verantwortlich war.

Kapitel III erweitert die Ergebnisse aus Kapitel II für normale Kinder auf Kinder mit klinisch beeinträchtigten arithmetischen Kompetenzen, d.h. einer Entwicklungsdyskalkulie. Entwicklungsdyskalkulie ist ein isoliertes Defizit in der Verarbeitung numerischen Materials und der Durchführung arithmetischer Operationen. Manche Kinder entwickeln eine Dyskalkulie, weil ihnen die grundlegende Vorstellung von Mengen aufgrund von genetischen Konditionen oder intrauterinen Komplikationen fehlt. Diese Kinder zeigen schwere Defizite in der Verarbeitung analoger Mengen sowie verbaler oder visueller Zahlennotationen. Andere Kinder jedoch entwickeln eine Dyskalkulie, weil ihre sprachliche Verarbeitung Defizite aufweist, die eine stabile Assoziation zwischen analogen Mengen und symbolischen Mengenrepräsentationen verhindern. Während diese Kinder ein Defizit bei der Enkodierung verbaler oder visueller Zahlenrepräsentationen haben, bleibt die Repräsentation analoger Mengen intakt. In der Studie lösten Kinder mit der bestätigten Diagnose einer Dyskalkulie Bayesianische Inferenzaufgaben und erhielten die Informationen entweder als Zahlensymbole (natürliche Häufigkeiten), oder als analoge Mengen. Kinder konnten mehr Aufgaben mit analogen Mengen lösen als mit natürlichen Häufigkeiten. Die anschließende Analyse zeigte, dass der positive Effekt analoger Mengen von der Unversehrtheit der analogen Mengenrepräsentation der Kinder abhängt. Während analoge Mengen in der Regel hilfreich für Kinder mit Dyskalkulie waren, konnten im Einklang mit klinischen Modellen der Dyskalkulie Kinder mit klinisch beeinträchtigten analogen Mengenrepräsentationen weniger aussagekräftige Informationen aus analogen Mengen extrahieren als Kindern, deren Beeinträchtigung auf symbolische Mengenrepräsentationen beschränkt war.

Schließlich untersucht Kapitel IV, ob die Unterschiede in der Anzahl der richtigen Lösungen, die bei Bayesianischen Inferenzaufgaben mit verschiedenen Repräsentationen beobachtet wurden, auch in Reaktionszeiten festgestellt werden können, und ob die verschiedenen arithmetischen Operationen die für die Berechnung von *a-priori*-Wahrscheinlichkeiten durchgeführt werden müssen, unterschiedliche Aspekte des Arbeitsgedächtnisses rekrutieren. Mit der Zahl der Rechenschritte steigt nicht nur die Wahrscheinlichkeit von Fehlern, sondern auch die Zeit, die es braucht, um die richtige Antwort zu ermitteln. Im ersten Teil dieses Kapitels wurden die Reaktionszeiten für Inferenzaufgaben mit verschiedenen Repräsentationen verglichen. Dabei ergab sich, dass, konform zur Komplexität der notwenigen arithmetischen Operationen, ein Abfall der Anzahl der richtigen Lösungen mit einem Anstieg der Reaktionszeit zusammenhing. Während analoge Mengen zu schnellen kategorialen Urteilen führten, war die Zeit, die benötigt wurde um *a-posteriori*-Wahrscheinlichkeiten zu ermitteln, abhängig von der Anzahl der einzelnen Elemente in der Menge. Natürliche Häufigkeiten zeigten schnelle Reaktionszeiten, sowohl für kategoriale Urteile als auch die Berechnung von Wahrscheinlichkeiten. Mit bedingten Wahrscheinlichkeiten benötigten die Probanden die meiste Zeit, unabhängig davon, ob sie nach kategorialen oder exakten Urteilen gefragt wurden. Nicht nur die Zahl der Rechenschritte unterscheidet sich zwischen der Repräsentationen, sondern die Art: Während natürliche Häufigkeiten eine Addition erfordern, sind mit bedingten Wahrscheinlichkeiten Multiplikationen notwendig, um zu korrekten Inferenzen zu gelangen. Analoge Mengen erfordern es, die einzelnen Ereignisse zu zählen, um *a-posteriori*-Wahrscheinlichkeiten zu berechnen. Additionen involvieren die Manipulation von Mengen auf einen mentalen Zahlenstrahl und rekrutieren visuell-räumliche Arbeitsgedächtnisprozesse. Multiplikationen hingegen benötigen verbal-auditive Arbeitsgedächtnisressourcen. Der zweite Teil des Kapitels IV zeigt die Ergebnisse einer Dual-Task-Studie, in der Teilnehmer Bayesianische Inferenzaufgaben mit unterschiedlichen Repräsentationen zeitgleich zu Reaktionsaufgaben durchführen mussten, die entweder Ressourcen des verbal-auditiven oder visuell-räumlichen Arbeitsgedächtnissystems blockierten. Wie angenommen wurde die Inferenzleistung mit natürlichen Häufigkeiten durch eine visuell-räumliche, nicht aber eine verbal-auditiven Zweitaufgabe vermindert, während das entgegengesetzte Muster für analoge Mengen und bedingte Wahrscheinlichkeiten gefunden wurde. Mit natürlichen Häufigkeiten und bedingten Wahrscheinlichkeiten waren auch kategoriale Inferenzen anfällig für den Einfluss einer Zweitaufgabe, während für analoge Mengen weder die Rate der richtigen Inferenzen noch die Reaktionszeiten beeinflusst wurden. Möglicherweise lässt sich die Bedeutung von Mengeninformationen aus analogen Mengen ohne Intention und den Einsatz von Arbeitsgedächtnisressourcen extrahieren.

Insgesamt unterstützt die in dieser Dissertation präsentierte Forschung beide Erklärungsansätze – dass die Darstellung von Informationen, wie sie in natürlichen Umgebungen vorkommen, Bayesianische Inferenzen erleichtert, und dass die Unterschiede in der Komplexität und Art der arithmetischen Operationen, die mit den unterschiedlichen Formaten verbunden sind, die Rate der richtigen Inferenzen und die Reaktionszeiten bestimmt. Ich hoffe, mit dieser Arbeit einige neue Elemente zu diesen Erklärungsansätzen hinzuzufügen. Zum einen konnte ich zeigen, dass auch Kinder in der zweiten Klasse systematisch Informationen nach dem Satz von Bayes integrieren können, wenn diese Informationen als Mengen dargeboten werden, die sie begreifen können. Zweitens sind auch Kinder mit schweren Beeinträchtigungen bei der Durchführung arithmetischer Operationen kleine Bayesianer, wenn die Informationen als analoge Mengen dargeboten werden. Dieser Effekt ist stärker für Kinder mit intakten Darstellungen von analogen Größen. Und schließlich werden in dieser Dissertation zum ersten Mal Reaktionszeiten für Bayesianische Inferenzaufgaben berichtet. Angesichts der langen Geschichte der Reaktionszeitanalyse als Mittel zur Erfassung der Komplexität mentaler

Operationen und der Differenzierung kognitiver Prozesse sollten sie zukünftig bei Studien zu kognitiven Prozessen bei Bayesianischen Inferenzen zusätzlich zu der Rate korrekter Antworten erhoben werden. Ebenso wurde zum ersten Mal ein Dual-Task-Ansatz verwendet, um die Anforderungen verschiedener Repräsentationen an Arbeitsgedächtnisprozesse in Bayesianischen Inferenzaufgaben zu untersuchen.

Die vorliegende Arbeit erweitert die bisherigen Erklärungsansätze, Erleichterung der Berechnung und ökologisches Design, und bietet Anknüpfungspunkte zu Theorien der allgemeinen und klinischen Entwicklungspsychologie, sowie zu Modellen des Arbeitsgedächtnisses. Ich hoffe diese Dissertation erweist sich als nützlicher Ausgangspunkt, um die Verbindungen zu vertiefen.

# Erklärung

Hiermit versichere ich, dass ich die vorgelegte Arbeit

*Representations Facilitate Bayesian Reasoning – Computational Facilitation and Ecological Design Revisited*

selbstständig verfasst habe. Ich habe keine anderen als die angegebenen Hilfsmittel verwendet. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Die Arbeit ist nicht als Ganzes oder in Teilen veröffentlicht. Teile der Daten aus Kapitel II sind bei Konferenzen vorgestellt worden (s. Curriculum Vitae). In der näheren Zukunft ist vorgesehen, die in Kapitel II vorgestellten Ergebnisse in überarbeiteter Form in Fachzeitschriften zu publizieren. Ebenso sollen die in Kapitel III berichteten Ergebnisse Ausgangspunkt für eine Fachpublikation sein. Koautoren werden in beiden Fällen Gerd Gigerenzer und Odette Wegwarth sein. Eine Publikation der Ergebnisse aus Kapitel IV in erweiterter und überarbeiteter Form ist mit Wolfgang Gaissmaier und Odette Wegwarth als Koautoren geplant.

Alle aufgeführten Koautoren werden bestätigen, dass ich hauptverantwortlich für das Entstehen der Kapitel war.

Jan Multmeier
Berlin, den 6. Februar 2012

# Curriculum Vitae

For reasons of data protection, the curriculum vitae is not included in the online version