Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Cognition 98 (2006) 287–308
COGNITION
www.elsevier.com/locate/COGNIT

Children can solve Bayesian problems: the role of representation in mental computation

Liqi Zhu[a], Gerd Gigerenzer[b,*]

[a]*Key Lab of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China*
[b]*Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany*

Received 15 July 2003; revised 26 July 2004; accepted 13 December 2004

Abstract

Can children reason the Bayesian way? We argue that the answer to this question depends on how numbers are represented, because a representation can do part of the computation. We test, for the first time, whether Bayesian reasoning can be elicited in children by means of natural frequencies. We show that when information was presented to fourth, fifth, and sixth graders in terms of probabilities, their ability to estimate the Bayesian posterior probability was zero. Yet when the same information was presented in natural frequencies, Bayesian reasoning showed a steady increase from fourth to sixth grade, reaching an average level of 19, 39, and 53%, respectively, in two studies. Sixth graders' performance with natural frequencies matched the performance of adults with probabilities. But this general increase was accompanied by striking individual differences. More than half of the sixth graders solved most or all problems, whereas one third could not solve a single one. An analysis of the children's responses provides evidence for the use of three non-Bayesian strategies. These follow an overlapping wave model of development and continue to be observed in the minds of adults. More so than adults' probabilistic reasoning, children's reasoning depends on a proper representation of information.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Bayesian problems; Computation; Binary hypothesis

* Corresponding author. Tel.: +49 30824 06 460.
E-mail address: gigerenzer@mpib-berlin.mpg.de (G. Gigerenzer).

0022-2860/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2004.12.003

There is an apparent paradox concerning the human ability to reason according to the laws of probability. On the one hand, according to the classical work by Piaget and Inhelder (1951/1975), by age 12 or so, children understand the laws of combinatorics, the law of large numbers, the irreversibility of chance processes, and other characteristics of probability. Recent studies have qualified this finding and demonstrated that, depending on the task, even younger children show signs of probabilistic reasoning (e.g. Falk & Wilkening, 1998; Schlottmann & Anderson, 1994). Piaget and Inhelder's work stands in the epistemological tradition of the Enlightenment's view of probability, according to which the laws of probability and the laws of human reasoning are two sides of the same coin (Daston, 1988). In Laplace's (1814/1951) famous phrase, the theory of probability is "only common sense reduced to calculus" (p. 196).

In apparent contradiction to these results, research in cognitive and social psychology since the 1970s has been interpreted as demonstrating that many adults fail to reason according to the laws of probability (e.g. Gilovich, Griffin, & Kahneman, 2002). In the cognitive illusions program, the very same laws that the Enlightenment probabilists and their modern followers, including Piaget, thought to match intuitive reasoning—such as the law of large numbers and set inclusion—were seen as difficult for human intuition. More recently, many of the original negative findings have been qualified and revised (e.g. Juslin, Winman, & Ohlson 2000; Kahneman & Frederick, 2002; Koehler, 1996a; Lopes, 1991). Nevertheless, a question arises. Why should Geneva children reason according to the laws of probability when Stanford undergraduates seem to be have problems doing so? This apparent paradox of early competence versus late failure has puzzled several scholars (e.g. Schlottmann, 2001).

In this article, we deal with a form of probabilistic reasoning that is considered difficult, namely, Bayesian reasoning. Do children reason the Bayesian way? To the best of our knowledge, there is no single study on children's Bayesian intuitions. There has been research with bees, bumblebees, birds, and other animals, concluding that animals are fairly good Bayesians (e.g. Real, 1991; Real & Caraco, 1986), whereas research with adults often came to the opposite conclusion. We will argue that the question "Do children reason the Bayesian way?" can only be answered when it is posed conditional to the external representation of numerical information.

1. Bayesian reasoning

Piaget and Inhelder (1951/1975) discussed the ideas of various mathematicians, mostly frequentists such as von Mises and Reichenbach, but Bayes's rule is never mentioned. In this paper, we deal with an elementary form of Bayesian reasoning, which has been the focus of almost all studies in adults (see Krauss, Martignon, & Hoffrage, 1999, for an exception). The elementary situation consists of a binary hypothesis (e.g. disease or no disease) and binary data (test positive or negative). The task is to evaluate the chances of the hypothesis given the data (e.g. of disease given a positive test). When we use the term "Bayesian reasoning" in this paper, we refer to reasoning in regard to this elementary situation.

For binary hypotheses (H and not-H) and data D, Bayes's rule is:

$$p(H/D) = \frac{p(H)p(D/H)}{p(H)p(D/H) + p(\text{not-}H)p(D/\text{not-}H)} \tag{1}$$

In words, Bayes's rule specifies how to derive the posterior probability $p(H|D)$ of a hypothesis with the provided data. This probability can be derived from the base rate (or prior probability) $p(H)$ in light of new data. The impact of the data is specified by the conditional probabilities $p(D|H)$ and $p(D|\text{not-}H)$. The rule is commonly attributed to Thomas Bayes, but its origin is uncertain. Stigler (1983) concluded that the odds are that Nicholas Saunderson, who held Newton's chair at Cambridge, actually discovered the rule.

2. The impact of representation on Bayesian reasoning

Conditional probabilities, as in Eq. (1), are a recent way to represent uncertainties, dating back to the invention of the mathematical theory of reasonableness in the mid-17th century (Gigerenzer et al., 1989). There are different ways of representing uncertainties. One representation is in terms of natural frequencies, which correspond to the result of observing outcomes in natural environments, that is, *counting without normalizing*. This has been the format in which humans—and animals—have encountered information in their environments during most of their history.

We illustrate the difference between a representation in terms of conditional probabilities and natural frequencies with one of the problems ("Red Nose") given to the children in this study.

Red nose problem: conditional probabilities

Pingping goes to a small village to ask for directions. In this village, the probability that the person he meets will lie is 10%. If a person lies, the probability that he/she has a red nose is 80%. If a person doesn't lie, the probability that he/she also has a red nose is 10%. Imagine that Pingping meets someone in the village with a red nose. What is the probability that the person will lie?

Red nose problem: natural frequencies

Pingping goes to a small village to ask for directions. In this village, 10 out of every 100 people will lie. Of the 10 people who lie, 8 have a red nose. Of the remaining 90 people who don't lie, 9 also have a red nose. Imagine that Pingping meets a group of people in the village with red noses. How many of these people will lie? ___ out of ____.

When the information is in conditional probabilities, the solution can be obtained by inserting the probabilities into Bayes's rule, as defined by Eq. (1). The result is

a probability of 47% that the person will lie. With natural frequencies, Bayesian computations become simpler (Gigerenzer & Hoffrage, 1995; Kleiter, 1994):

$$p(H/D) = \frac{d}{d+f} \tag{2}$$

Natural frequencies are illustrated by the tree in Fig. 1. There are *a* cases (e.g. people), of which *b* cases show *H* (e.g. they are liars) and *c* cases do not. Among the *H* cases, there are *d* cases that show *D* (e.g. a cue such as red nose) and *e* cases that do not; among the not-*H* cases, *f* show *D* and *g* do not. Thus, the answer is: Eight out of 17 people with a red nose will lie, which corresponds to a probability of 47%.

The reason for the computational simplification is that natural frequencies still contain information about base rates, whereas conditional probabilities are obtained by taking the base rate information out of the natural frequencies (i.e. normalization). As a result, when one transforms the natural frequencies in Eq. (2) into probabilities, the base rates have to be put back into Eq. (1) by multiplying the conditional probabilities by the base rates. This can be seen by comparing the structures of Eqs. (1) and (2), which are exactly the same.

Note that the natural frequencies correspond to the way a child would learn from direct observation, whereas conditional probabilities are derived by normalizing natural frequencies. It is important to keep in mind that not all frequencies are natural frequencies. For instance, relative frequencies do not lead to computational simplification (Gigerenzer & Hoffrage, 1995, Prediction 4 and Experiment 2; 1999). Some researchers, however, have confused natural frequencies with *any* kind of frequencies. For instance, Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni (1999) stated that "in fact, data in the form of frequencies by no means guarantee good Bayesian reasoning" (p. 81), and referred to a study in which *normalized*, but *not natural frequencies* were provided. Similarly, Evans, Handley, Perham, Over, and Thompson (2000) tested normalized frequencies, found no difference from conditional probabilities, concluded that frequencies per se do not explain the facilitation of Bayesian reasoning, and presented

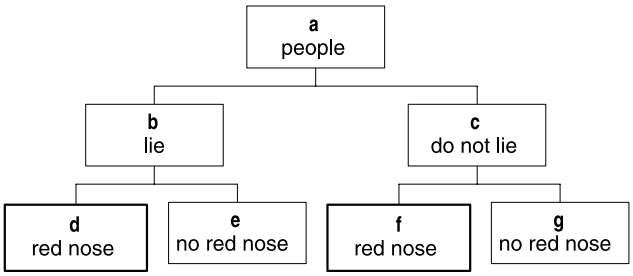

Fig. 1. Illustration of a natural frequency tree. The total observed frequency *a* is split into *b* (e.g. the number of people who lie) and *c* (the number of people who do not lie). The frequency *b* out of *a* is the base rate (e.g. of lying). The frequency *b* is, in turn, split into *d* (e.g. the number of people with red noses among those who lie) and *e* (the number of people without red noses among those who lie). The frequency *d* out of *b* is the hit rate (or sensitivity). Similarly, the frequency *c* is split into *f* (e.g. the number of people with red noses among those who do not lie) and *g* (the number of people without red noses among those who do not lie). The frequency *f* out of *c* is the false positive rate. Bayes's rule amounts to $d/(d+f)$, that is, only the two natural frequencies in the bolded boxes need to be attended to. The base rate can be ignored because it is contained in these natural frequencies.

this result as evidence against Gigerenzer and Hoffrage's (1995, 1999) thesis. Yet the argument is that representations do part of the computations, not that frequencies per se would, for some unknown reason, facilitate Bayesian computations (see Brase, 2002; Hoffrage, Gigerenzer, Krauss, & Martignon, 2002).

3. Multiple cognitive strategies

How do children and adults try to solve Bayesian-type tasks such as the Red Nose problem? Previous research on children's cognitive strategies in different tasks suggests that there is often not a single strategy (such as an adding rule), but rather a toolbox of strategies available (e.g. Siegler, 1999, 2000). Similarly, research on Bayesian tasks has shown that adults use about five different strategies, and that the prevalence of these strategies varies with the format of information (Gigerenzer & Hoffrage, 1995; Hoffrage & Gigerenzer, 1998). The possible existence of multiple strategies in children has methodological consequences. First, we will analyze strategies of children at the individual level, rather than at the level of aggregates of individuals. Second, given the observation in previous research with adults that individuals sometimes switch between strategies from problem to problem rather than consistently using only one strategy, we will also analyze an individual's strategy for each problem rather than aggregating across all problems. In other words, the unit of analysis is an individual-task combination. Finally, the existence of multiple strategies requires testing the data of each individual-task combination against multiple hypotheses (i.e. the candidate strategies), rather than one null hypothesis.

The frequency distributions of cognitive strategies can change systematically over time. Strategies that are frequent in early age often drop out over time, and new strategies join and may become dominant. For instance, Siegler (1999) and Shrager and Siegler (1998) described the developmental change in strategies that preschoolers use for addition, such as $3+5$, and Gigerenzer and Richter (1990) analyzed the developmental change in strategies used to estimate the area of rectangular shapes. The common observation is that strategy development does not proceed from using one strategy at one stage to using a different strategy at a second stage, and so on. Rather, developmental change seems to follow Siegler's (1999) overlapping waves model, in which multiple strategies coexist at each point; what changes is their onset and prevalence. The methodological consequence of this observation is again to analyze strategies at the individual (and task) level, and to describe the rise and fall of strategies throughout development in terms of the frequency distribution of each strategy.

4. Questions

This article addresses three questions:

1. Does children's ability to solve Bayesian problems depend on the information presentation? Specifically, do natural frequencies help children to reason the Bayesian way? If so, at what age?

2. What are the cognitive strategies children use to solve Bayesian problems? How do they differ from the strategies adults use?
3. What is the developmental pattern of change in strategies?

We performed two studies with a similar design. Because both studies had the same structure, we will report the results together.

5. Method

5.1. Participants, age, and design

In both studies, the children were from ordinary elementary schools in Beijing. The adults were MBA students from the School of Management, Beijing University of Aeronautics and Astronautics. In Study 1, the mean age of the fourth graders was 9;9 (range from 9;4 to 10;3), the mean age of the fifth graders was 10;8 (range from 10;1 to 11;4), and the mean age of the sixth graders was 11;9 (range from 11;3 to 12;6). The mean age of the adults was 29 (range from 25 to 35). The number of participants was 16, 15, 14, 23, and 23, for the fourth, fifth, and sixth graders, and the two adult groups, respectively. In Study 1, children were tested with natural frequencies alone, whereas one group of adults was tested on conditional probabilities and the other on natural frequencies.

In Study 2, the mean age of the fourth graders was 9;7 (range from 9;2 to 10;1), the mean age of the fifth graders was 10;7 (range from 10;1 to 11;3), and the mean age of the sixth graders was 11;6 (range from 11;2 to 12;2). The mean age of the adults was 20 (range from 19 to 23). The number of participants was 30 in each of the four groups for the natural frequency representation. In addition, we tested children and adults on the probability version, with 10 children in each of the three age groups and a total of 30 adults**.** We used a smaller number of children for the probability condition because we feared that establishing the probability base line might frustrate them if the task were too difficult.

5.2. Reasoning problems

We constructed 10 Bayesian problems whose content was suited to children. One of these is the Red Nose problem described above; the others are given in the Appendix. Study 1 used 7 and Study 2 all 10 problems. As the Red Nose problem shows, probabilities were always expressed as percentages.

5.3. Procedure

Children were tested in small groups of three to six persons. The adults were tested in larger groups in a classroom. In Study 2, children were randomly assigned to either conditional probability or natural frequency representation, and thus the factor representation was varied between participants. In Study 1, children were only tested on natural frequencies. The instruction was the same for all age groups: “Please solve

the following problems. Each problem includes several numbers and a question. Please write down how you got the answer or mark the numbers you used to get the answer."

5.4. Criteria for identifying Bayesian responses

When the Bayesian answer is 8 out of 17 (or a probability of 47%), should a child's answer "9 out of 17" (or a probability of 53%) count as a Bayesian response? To avoid classifying an answer as a Bayesian response although the underlying strategy was in fact not a Bayesian strategy, we used a strict outcome criterion (similar to Gigerenzer & Hoffrage, 1995): An individual's response has to be numerically exactly the same as the Bayesian solution. By the phrase "exactly the same," we mean that children and adults have to report the exact frequency or probability; otherwise it is not coded as a Bayesian response. For instance, in the Red Nose problem, the answer "9 out of 17" was not coded as a Bayesian response, although the absolute difference from "8 out of 17" may be considered an insignificantly small deviation in a significance test. The strict criterion diminishes the probability that a mere guess is mistakenly classified as a Bayesian response (see below).

In addition, participants were encouraged to circle the numbers they had used. For instance, one child wrote "$21+28=49$, $21/49=3/7$", circled the "21" and the "28," and answered "3 out of 7" (for Problem 2, see Appendix). The circled numbers were used as a double check: If the circled numbers were inconsistent with Bayesian reasoning, then the response was not counted as a Bayesian response. We did not find such cases; children's numerical answers and the numbers they circled were consistent.

The opposite error, not classifying Bayesian reasoning as a Bayesian response, might occur when a person reasons the Bayesian way but makes an error in the calculation, for instance, when the information is in probabilities that involve substantial calculations. Therefore, if a person showed Bayesian reasoning in terms of process—for instance, when writing down the equation $0.10\times0.80/(0.10\times0.80+0.90\times0.10)$ for the Red Nose problem—but subsequently made a calculation error, or did not calculate the result, we classified this as a Bayesian solution. The reason is that we are investigating people's Bayesian reasoning, not their calculation skills. These cases occurred only with probability formats, and only in adults (in 12 out of a total of 230 Bayesian responses).

6. Results

6.1. Can children reason the Bayesian way with conditional probabilities?

The results with probability format were straightforward (Fig. 2). Adults were able to solve 49% of the problems (and 57% in Study 1, where children were not tested with a probability format). None of the fourth graders, none of the fifth graders, and none of the sixth graders could solve any of the problems when the information was in probabilities. Their strategies seemed to be random: They picked one of the percentages, or added or subtracted two or three percentages, and sometimes answered that the probability was 160% or a similarly impossible value. The children seemed to have no clue how to solve

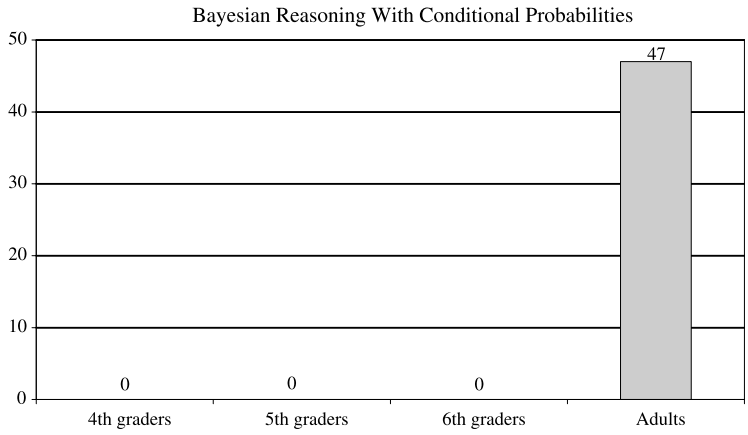

Fig. 2. Percentage of Bayesian solutions with conditional probabilities. Based on 60 participants (10 in each of the children groups and 30 adults) and 10 problems, there were a total of 600 individual problems. None of the children could solve any of the problems.

the problems when the information was in terms of probabilities. We take this result as a base line for the possible effect of natural frequencies.

The inability of children to solve these problems should be evaluated against the fact that they were not familiar with the mathematical concepts of probability and percentage. Percentages were, at the time of the study, not taught in Chinese primary schools before the middle of sixth grade (the sixth graders in this study were tested before they were exposed to percentages). The results in Fig. 2 were consistent with the expectations of the children's teachers, who predicted that the children would not be able to solve these problems, given that they had never been exposed to them. Were the teachers' expectations still true when probabilities were replaced by natural frequencies?

6.2. Can natural frequencies improve Bayesian reasoning in children?

Fourth graders were able to solve 17 and 19% of the problems in Study 1 and 2, respectively, when the information was in natural frequencies (Fig. 3). For the fifth graders, this number increased to 30 and 42%. By sixth grade, children solved 70 and 48% of the problems. Adults reached a performance of 75 and 77% in Study 1 and 2, respectively. Averaged across both studies and weighted by the sample sizes, the results are 18.7, 39.0, 53.5, and 76.1% Bayesian solutions for the fourth, fifth, and six graders, and the adults, respectively. When one compares across representations, the sixth graders' performance with natural frequencies matched and surpassed the performance of adults with probabilities (Fig. 2).

Table 1 shows that the Bayesian responses were not uniformly distributed across participants, but showed marked individual differences. Among the fourth graders, a large proportion could not solve a single one of the problems. Most others succeeded with one or two of the problems. These could be signs of an occasional flash of Bayesian reasoning in

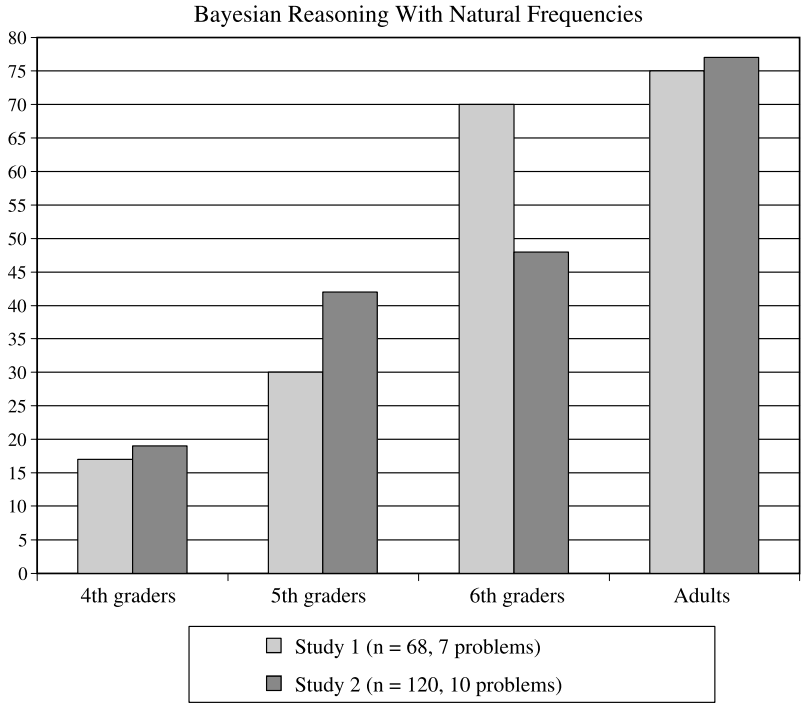

Fig. 3. Percentage of Bayesian solutions with natural frequencies. In Study 1, there were 16, 15, 14, and 23 participants in the four age groups, respectively. In Study 2, there were 30 participants in each group. The total number of individual problems (natural frequency representations) was 476 in Study 1, and 1200 in Study 2.

this age group. Most remarkably, in Study 2, one child solved all 10 problems and three solved 8 or 9 of the problems. These extraordinary children reasoned systematically better—with natural frequencies—than a number of the adult MBA students.

The picture for the fifth graders is different. The proportion of Bayesian responses was twice as high (Fig. 2) and the interindividual variability increased (Table 1). In Study 1, 5 out of the 15 children gave the Bayesian response most of the time, whereas 6 children showed no sign of Bayesian reasoning—the same number as among the fourth graders. In Study 2, 7 out of 30 children found the Bayesian solution for every problem, whereas 13 could not solve a single problem.

For the sixth graders, there were essentially two groups left. The majority of children now reasoned the Bayesian way for most or all of the problems. At the same time, almost one third of the sixth graders still showed no sign of Bayesian intuition and could solve none of the problems.

To summarize: At an age where conditional probabilities pose immense difficulties, natural frequencies can foster Bayesian reasoning in children. Sixth graders' performance with natural frequencies matched the performance of adults with probabilities. But the general increase in performance was accompanied by striking individual differences.

Table 1
Distribution of the number of Bayesian responses

Grade	Number of Bayesian responses: Study 1								*n*	Total
	0	1	2	3	4	5	6	7		
Fourth	6	2	7	1	0	0	0	0	16	19
Fifth	6	2	1	1	2	2	1	0	15	31
Sixth	3	0	0	1	0	0	4	6	14	69
Adults	4	1	0	0	1	0	3	14	23	121

Grade	Number of Bayesian responses: Study 2											*n*	Total
	0	1	2	3	4	5	6	7	8	9	10		
Fourth	18	5	0	0	0	1	1	1	2	1	1	30	58
Fifth	13	1	0	2	1	0	2	0	2	2	7	30	127
Sixth	11	1	2	0	0	0	0	1	8	2	5	30	144
Adults	5	0	1	0	0	0	2	0	0	4	18	30	230

n, Number of participants in each group; Total, number of Bayesian responses in each group. For each participant, the total number of Bayesian responses could vary between 0 and 7 in Study 1 (top panel) and between 0 and 10 in Study 2 (bottom panel). For instance, in Study 1, six children in fourth grade and four adults could not solve any of the problems.

Whereas more than half of the sixth graders solved most or all problems, most of the others could not solve a single one.

How do children (and adults) who do not reason the Bayesian way attempt to solve the problems? Do they merely guess or do they use systematic, pre-Bayesian strategies?

6.3. Non-Bayesian intuitions

In the absence of previous research with children, we started with the three most frequent non-Bayesian strategies in adults as hypotheses (Gigerenzer & Hoffrage, 1995). The *joint occurrence* strategy is related to the positive testing strategy (Klayman & Ha, 1987), where people only look for the frequency of confirming evidence, such as when both symptom and disease occur together. For this and other non-Bayesian strategies, we used the same criteria as for identifying Bayesian reasoning. In the Red Nose problem, the joint occurrence strategy predicts that children will use the number of people who have a red nose and lie (here: 8 out of 100). However, the children in our two studies never followed this strategy. Instead, we observed cases consistent with the two other strategies adults use, *conservatism* and *representative thinking*, along with two strategies that have not been reported in earlier studies. The logic of the resulting four strategies is shown in Fig. 4, using the Red Nose problem.

If a child centers on only one aspect of the problem, then this aspect is either the event to be predicted (e.g. people who lie), or the evidence available (e.g. red nose). If the child centers on the event, the following strategy results (*a* and *b* refer to Fig. 1).

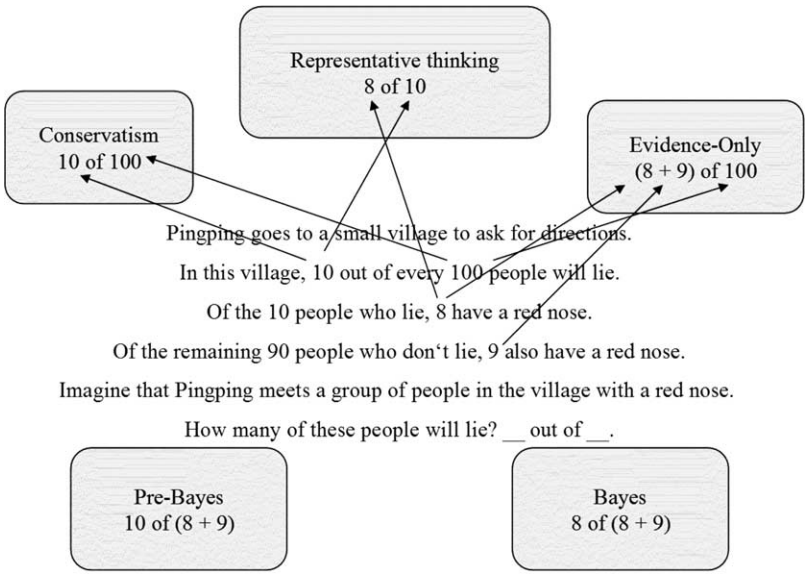

Fig. 4. Non-Bayesian and Bayesian strategies in children. For three of the strategies, arrows show what information children pay attention to (to avoid illegibility, arrows are omitted for two strategies). The formula underneath the name shows how the numbers are combined. For instance, the strategy called conservatism leads to the answer that "10 out of 100 people will lie.".

Conservatism: b/a

The underlying intuition is to stick with one's prior beliefs and disregard new evidence. This strategy has been labeled "conservatism" (Edwards, 1968) and "base-rate only" (Gigerenzer & Hoffrage, 1995). It amounts to the opposite of base rate neglect. In the Red Nose problem, conservatism generates the answer "10 out of 100" rather than "8 out of 17." It typically underestimates the Bayesian probability or frequency. In studies with natural frequencies, a small number of adult laypeople (Gigerenzer & Hoffrage, 1995) and physicians making diagnostic inferences (Hoffrage & Gigerenzer, 1998) have been reported to rely on conservatism.

If the child centers only on the evidence, then a different strategy results.

Evidence-only: (d+f)/a

The intuition underlying this strategy is the opposite of conservatism: People jump at the new evidence and disregard prior beliefs. Evidence-only disregards the base rate of the target event, and focuses only on how often the evidence or signal ($d+f$) occurred among all cases (a). In the Red Nose problem, evidence-only generates the answer "17 out of 100." This strategy was not reported by Gigerenzer and Hoffrage (1995). The third strategy has been observed frequently in adults and uses a different denominator than the two centering strategies.

Representative thinking: d/b.

The intuition underlying this strategy is to see how often the evidence (e.g. red nose) occurs when the event (lying) is there. This strategy has been variously called "representative thinking" (Dawes, 1986), or "Fisherian" (Gigerenzer & Hoffrage, 1995). The reason for the latter is because it corresponds to a widely-used method of hypothesis testing, known as significance testing. In Fisher's theory of significance testing, an inference from data D to a null hypothesis H_o is based solely on $p(D|H_o)$, which is known as the exact level of significance. As Fig. 1 illustrates, the proportion d of b is equivalent to $p(D|H)$. Fisher was not a Bayesian; in fact, he ridiculed Bayesian statistics as useless in science (Gigerenzer et al., 1989, chap. 3). Unlike Bayes's rule, significance testing ignores both base rates and false positive rates. Representative thinking produces the confusion of the Bayesian posterior probability with the hit rate, and has been reported to describe some physicians' and lawyers' intuitions when the information is presented in conditional probabilities (Dawes, 1986; Koehler, 1996b). In the Red Nose problem, this strategy generates the response "8 out of 10" rather than the correct answer "8 out of 17."

Only one strategy had the same denominator as Bayes's rule.

Pre-Bayes: b/(d+f)

To the best of our knowledge, this strategy has not been reported in the literature before. We call it pre-Bayes because it gets the denominator right and the numerator approximates the numerator in Bayes's rule for high hit rates. Children who use this strategy focus on the number b of target events (such as cases of lying) rather than on the number d of cases where a cue (red nose) is present among the target events. In the Red Nose problem, this strategy generates the answer "10 out of 17" instead of the correct answer "8 out of 17." As this example illustrates, picking b instead of d can lead to estimates close to answers based on Bayesian reasoning.

Some children wrote down clearly how they arrived at their answer, and here we also have process information for the non-Bayesian strategies. For instance, for problem 2 (salty cookies), one child wrote "20/100 = 1/5", and then responded "1 out of 5" (conservatism). Another child wrote "14 + 24 = 38, 38:100 = 38/100", and then answered "38 out of 100" (evidence-only); and a third child wrote "14 + 24 = 38, 20:38 = 20/38 = 10/19", and then answered "10 out of 19" (pre-Bayes).

Tables 2 and 3 show how often children and adults followed each of the four strategies. For instance, in Study 1, three fourth graders followed pre-Bayes in two out of seven problems, and one fifth grader did so consistently in all seven problems.

But how can we know whether these response patterns are the product of systematic strategies, rather than of mere guessing by picking numbers randomly? Specifically, we need to ask for each candidate strategy (a) whether there were more correct responses in total than would have been expected if the children had guessed, and (b) whether some children followed the strategy systematically. It is not easy to define the set of options from which children could randomly pick. We suggest doing this on the basis of two observations: (i) Children consistently picked either two or three but not more of the total of five numbers specified in each problem (e.g. two for conservatism and representative

Table 2
Distribution of the Number of Children's responses following each of four strategies in Study 1

Grade	Pre-Bayes							total
	1	2	3	4	5	6	7	
Fourth	4	3	0	0	0	0	0	10
Fifth	2	1	1	0	2	0	1	24
Sixth	6	0	0	0	0	0	0	6
Adults	0	0	0	0	0	0	0	0

Grade	Conservatism							total
	1	2	3	4	5	6	7	
Fourth	5	5	0	0	0	1	0	21
Fifth	2	0	0	0	1	0	1	14
Sixth	0	0	0	0	0	0	0	0
Adults	1	0	0	0	0	1	0	7

Grade	Evidence-only							total
	1	2	3	4	5	6	7	
Fourth	2	3	0	0	0	0	0	8
Fifth	0	0	1	1	1	0	0	12
Sixth	0	0	0	0	0	0	0	0
Adults	0	0	0	0	0	0	0	0

Grade	Representative thinking							total
	1	2	3	4	5	6	7	
Fourth	6	3	0	0	0	0	0	12
Fifth	0	0	0	0	0	0	0	0
Sixth	0	0	0	0	0	0	0	0
Adults	1	0	0	0	0	0	0	1

The shaded part indicates response patterns whose probability under the null hypothesis that a child picked numbers randomly is $P = 0.0005$ or smaller (see text). The sample size is $n = 16$, 15, 14, and 23 in fourth, fifth, and sixth graders, and adults, respectively.

thinking, three for pre-Bayes and evidence-only), and (ii) when children picked three numbers, they added—rather than subtracted—two of them. This results in 20 ways to pick two out of five numbers, and in 60 ways to pick three. These 80 random choices reduce to 40 because of the logical constraint that the second number in the response "__ out of __" cannot be smaller than the first, a constraint that children consistently followed.[1] For simplicity, we define the chance hypothesis by a uniform distribution, that is, random picking means that each of these 40 patterns has the same probability.

The first question is whether the total number of patterns consistent with each of the four candidate strategies is actually larger than what can be expected by chance. The total

[1] One can generate 5×4 ordered pairs from five numbers, and one of each pair (a, b) and (b, a) is eliminated by the logical constraint. This coincides with the number of nonordered pairs (where the order of a and b is irrelevant), which amounts to 10. Similarly, one can generate 10 nonordered triples out of 5 numbers. Each triple (a, b, c) of numbers picked can result in six judgments, "a out of b + c", "b out of a + c", "c out of a + b", "a + b out of c", "a + c out of b", and "b + c out of a." Since there are ten triples, this amounts to 60 judgments. The logical constraints reduces these by half.

Table 3
Distribution of the number of children's responses following each of four strategies in Study 2

	Pre- Bayes										
Grade	1	2	3	4	5	6	7	8	9	10	total
Fourth	4	2	0	0	0	1	0	0	0	0	14
Fifth	0	1	0	1	0	0	2	0	0	0	20
Sixth	4	3	1	0	0	1	1	5	0	0	66
Adults	2	0	1	0	0	0	1	0	0	1	22

	Conservatism										
Grade	1	2	3	4	5	6	7	8	9	10	total
Fourth	4	1	2	1	0	0	0	0	0	0	16
Fifth	1	0	0	0	0	0	0	0	0	0	1
Sixth	1	1	0	0	0	0	0	0	0	1	13
Adults	1	0	0	0	1	0	0	0	0	0	6

	Evidence-only										
Grade	1	2	3	4	5	6	7	8	9	10	total
Fourth	0	1	0	0	0	0	0	0	0	0	2
Fifth	0	0	0	0	0	0	0	0	0	0	0
Sixth	0	0	1	0	1	1	0	0	0	2	34
Adults	0	0	0	0	0	0	0	0	0	0	0

	Representative thinking										
Grade	1	2	3	4	5	6	7	8	9	10	total
Fourth	1	0	2	0	0	0	0	0	0	0	7
Fifth	0	0	1	0	0	0	0	0	0	0	3
Sixth	0	0	0	0	0	0	0	0	0	0	0
Adults	0	0	0	0	0	0	0	0	0	0	0

The shaded part indicates response patterns whose probability under the null hypothesis that a child picked numbers randomly is $P=0.0005$ or smaller (see text). The sample size is $n=30$ in each group.

number of children's Bayesian responses is 119 (out of 315) in Study 1 and 329 (out of 900) in Study 2. Thus, we have 196 and 571 unaccounted responses, which total 767. Assuming that these children picked randomly, one expects $767/40=19.2$ picks that look like one of the four strategies, whereas the actual numbers were 140, 65, 56, and 22, for pre-Bayes, conservatism, evidence-only, and representative thinking, respectively (Tables 2 and 3). That is, patterns implied by the first three strategies occur much more frequently than expected by the chance model, whereas the same analysis shows no support for representative thinking.

This aggregate analysis, however, does not consider the possibility of systematic individual differences. Are there children who follow one strategy consistently? Consider first pre-Bayes (Tables 2 and 3). For Study 1, a binomial test ($P=1/40$; $n=7$) shows that the probability of producing three (or more) times the predicted pattern by chance is $P=0.0005$. Thus, three or more answers are unlikely to occur by chance. For the two fifth graders who followed the predicted pattern five out of seven times, the corresponding probability is in the order of one in a billion, and that of the fifth

grader who always followed the prediction of pre-Bayes is even smaller. For Study 2, a binomial test ($P = 1/40$; $n = 10$) indicates that similar systematic patterns of individual differences exist. Eleven children, for instance, followed pre-Bayes four or more times; in each case, the probability of this result under the chance hypothesis is $P = 0.00007$. The corresponding probability for the five sixth graders whose response pattern is consistent with the pre-Bayes pattern in 8 out of 10 problems is smaller than one in a billion. Thus, we can conclude that there is evidence that the responses consistent with pre-Bayes are not random but systematic. This strategy is most pronounced in grades 5 and 6.

By the same analysis, one can detect a comparatively small number of children who systematically follow conservatism, with two following this strategy in every problem. The same holds for the Evidence-Only strategy. In contrast, there is again relatively little support for representative thinking. Study 1 does not provide convincing evidence. In Study 2, there are three children whose responses are consistent with this strategy three times. The probability of this result under the random picking hypothesis is $P = 0.0005$ in every case. Nevertheless, we conclude that this candidate strategy does not have the magnitude of support that the other three strategies have; there is no single child who follows it most or all of the time. The support for representative thinking is therefore open, and requires further backing from independent studies.

In summary, when information is represented in terms of natural frequencies, Bayesian reasoning, as well as three non-Bayesian strategies, can be documented in children. One of them, pre-Bayes, is close to Bayesian reasoning. The others can lead to systematic and large deviations, because they center on only one aspect of the problem.

6.4. Is there a developmental change in strategies?

The main developmental change documented is the fast increase of Bayesian responses from approximately ages 10–12 (Fig. 3) when the information is in natural frequencies. A second major change concerned the rate of "guessing" (responses that were neither Bayesian nor one of the non-Bayesian strategies) and "no answer." Guessing decreased from 50% in fourth graders to 41, 12.6, and 16.1% in fifth graders, sixth graders, and adults, respectively. The percentage of participants who gave no answer decreased from 9.5% in fourth graders to 1.7, 4, and 0%, respectively.

Is there also a developmental trajectory in the non-Bayesian strategies? Fig. 5 suggests a developmental change akin to Siegler's overlapping waves model. There seem to be four waves:

Wave 1: *Guessing*. In this mental state of confusion, children and adults find neither the Bayesian answer nor a consistent strategy. The responses to the problems are based on arbitrarily picking a few numbers, or no answer. The first wave strongly decreases from an initial 50% in fourth graders, but never completely disappears and remains present in adult life.

Wave 2: *Centering*. Children center on one of the two aspects of the problem, the event or the evidence. This wave consists of strategies that have not yet found the proper denominator: conservatism and evidence-only. The developmental trajectory of

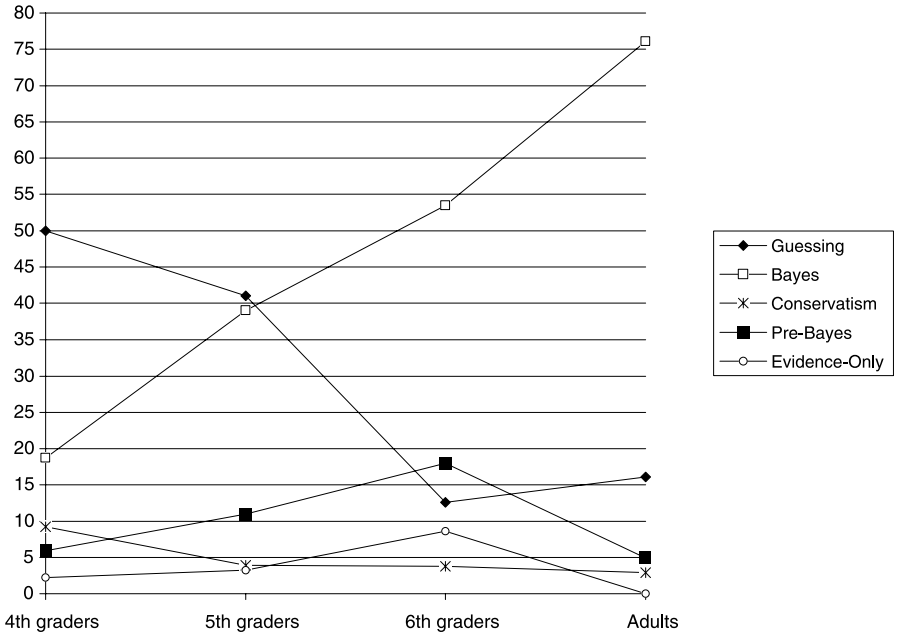

Fig. 5. Percentage of pre-Bayesian and Bayesian strategies across both studies. For instance, fourth graders tried to solve 9% of all problems by using the conservatism strategy.

the second wave is less clear than that of the first. The second wave is weaker than the first at all ages, and seems to have a slowly decreasing trajectory. It also extends into adult life.

Wave 3: *Pre-Bayes*. The third wave consists of a first approximization of Bayesian reasoning, where children no longer center on one aspect of the problem. They have found the proper denominator but not yet the numerator. The temporal overlap between the second and third wave seems substantial. Pre-Bayes peaks in the older children, and can be seen as a forerunner of Bayesian reasoning.

Wave 4: *Bayesian reasoning*. The fourth and last wave is Bayesian reasoning. Here, the pattern is again very clear. Bayesian reasoning increases monotonically with age and reaches its mode in adults.

We would like to emphasize that this sequence of waves should be seen as a hypothesis, for the data on which it is based is limited. Data on both younger and older children are required for further test.

7. Conclusion

This article provided an ecological perspective on the development of reasoning: The external representation does part of the internal computation. The major result is that children can systematically reason the Bayesian way if the information is provided in

natural frequencies rather than in probabilities. This does not exclude the possibility that other representations also foster insight, specifically if they mimic the structure of natural frequencies and lead to the same computational facilitation (e.g. Gigerenzer, 2002). More generally, any numerical information can be represented in various forms, as Roman and Arabic numbers illustrate, but these are not neutral forms for the same content, because they actually facilitate certain computations and insights, and hinder others (Martignon, Vitouch, Takezawa, & Forster, 2003). As the physicist Richard Feynman remarked, different representations of the same mathematical formula can evoke varied mental pictures and lead to new solutions (Feynman, 1967, p. 53).

Aside from doing part of the computation, representations influence thinking in another way. Representations can constrain the possible set of mental strategies, and thereby the resulting judgments. For instance, with Bayesian problems in the form of conditional probabilities, base rates need to be attended to, such as the base rate b/a of liars in Fig. 1. With natural frequencies, in contrast, base rates need not be attended to; the ratio b/a can be ignored. The reason is that base rate information is contained in natural frequencies, whereas conditional probabilities are normalized with respect to the base rates. Thus, when moving from natural frequency to other representations, one needs to learn to pay attention to base rates. Similarly, representations constrain the possible set of non-Bayesian strategies as well as Bayesian shortcuts; one cannot use every strategy with every representation (Gigerenzer & Hoffrage, 1995).

Natural frequency representations can provide explanations for other phenomena besides Bayesian reasoning. Consider an error commonly made by children (Carpenter, Coburn, Reys, & Wilson, 1978). When asked to find the sum of 1/2 and 1/3, the answer is often 2/5. This is called the freshman error of adding numerators and adding denominators (Silver, 1986). Natural frequency representations can offer an account for this error. If one assumes that natural frequencies such as "1 out of 2" and "1 out of 3" are developmentally primary and that relative frequencies (or fractions) such as 1/2 and 1/3 are only understood later in development, the error can be deduced. The reason is that one can add "numerators" and "denominators" of natural frequencies. For instance, in the Red Nose problem, "8 out of 10" people who lie have a red nose, and "9 out of 90" people who do not lie also have a red nose. You can add these to total "17 out of 100" people who have a red nose. This operation is correct in an environment with natural frequencies. In contrast, a relative frequency 8/10 is the same as 4/5, and adding the numerators and denominators of 4/5 and 9/90 will result in an error. Our explanation for the freshman error is that children or adults apply the principles of natural frequencies to fractions or relative frequencies.

Natural frequencies facilitate Bayesian reasoning because they work by simple enumeration without normalization (compare Eqs. (1) and (2)). The power of enumeration extends beyond the problems studied in this article. Infants are sensitive to changes in the numerosity of a collection of visual objects (Antell & Keating, 1983; Starkey & Cooper, 1980); at 6 months, they seem to form arithmetical expectations when an object is added or taken away from a small collection (Wynn, 1992). Fractions are hard to understand whereas collections seem to be easy (Brase, Cosmides, & Tooby, 1998), as illustrated by the almost universal use of fingers as a representational

system (Butterworth, 1999, 2001), and brain imaging studies suggest that key number areas are closely connected to the finger circuit (Dehaene, Spelke, Pinel, Stanescu, & Tviskin, 1999). Krauss and Wang (2003) studied the notorious Monty Hall Problem, which most adults fail at, and showed that correct answers can be facilitated by changing the single-event question ("should I switch or not switch doors") into a frequency question ("in how many cases will switching win?"), and Krauss and Atmaca (2004) showed similar facilitation for children and adolescents aged 11–19.

In this article, we asked: Can fourth, fifth, and sixth graders solve Bayesian problems when the information is in conditional probabilities? The answer is no. Can Bayesian reasoning be elicited when the representation is switched to natural frequencies? The answer is yes. By sixth grade, the effect of representation was as strong as or stronger than the average effect on university students, 16 versus 46% (15 problems, Gigerenzer & Hoffrage, 1995); on physicians, 10 versus 46% (four problems, Hoffrage & Gigerenzer, 1998); and on members of the National Academy of Neuropsychology, 9 versus 63% (one problem, Labarge, McCaffrey, & Brown, 2003; see also Betsch, Biel, Eddelbuttel, & Mock, 1998; Cosmides & Tooby, 1996; Gigerenzer, 2002; Hoffrage, Lindsay, Hertwig, & Gigerenzer, 2000).

Can the present results be generalized to children in other cultures with other educational backgrounds? In a replicate study of the present experiments with German children, Lücking (2004) has investigated fifth, sixth, and seventh graders. She chose children who were one year older because, just like the Chinese teachers, the German teachers and experts she consulted did not believe that children could solve these problems—whatever the external representation might be. Furthermore, given the higher mathematics performance and motivation of Chinese children compared to German children (Artelt et al., 2001; Stern, Rode, Fang, & Zhu, 2001), Lücking expected that German children would not be able to solve the problems as early as Chinese children. However, the performance of the German fifth and sixth graders was indistinguishable from that of the Chinese ones. She also tested the conjecture by Girotto and Gonzales (2001) that "single-step questions" (as used in the present study) would result in substantially lower numbers of correct solutions than "two-step questions" (where first the value of the denominator in Eq. (2), and then that of the numerator is asked), but found no evidence. Nor did additional questions that amplified the set relations between frequencies increase performance.

Let us return at the end to the beginning, to the Enlightenment's vision that human intuition and the laws of probability are just two sides of the same coin, and to the contrasting view of much of recent cognitive psychology. What does children's performance tell us about this apparent paradox? It offers a way to resolve the contradiction. The solution may lie in the representation, that is, in the way one presents information to children and adults. When one chooses a representation that comes natural to the human mind and reduces computation, the Enlightenment's view is not too far removed from the evidence. Confronted with conditional probabilities, in contrast, children are helpless, as are many adults. These results have implication for teaching statistical reasoning (Sedlmeier, 1999). Learning to play with representations should be part of mathematics education. To solve a problem involves finding a proper representation.

Acknowledgements

This research is supported in part by a grant from the Chinese Academy of Sciences (KSCX2-SW-221, 70471060).

Appendix A. Reasoning problems

The first 7 problems were used in Study 1; all 10 problems were used in Study 2. Problem 1 (Red Nose) is given in the text, both in natural frequencies and in probabilities. Problems 2 through 10 are given below in terms of natural frequencies. The probability version can be deduced from it, and is therefore omitted here.

2. There is a large package of sweet or salty cookies with various kinds of shapes. In the package, 20 out of every 100 cookies are salty. Of the 20 salty cookies, 14 are round. Of the remaining 80 sweet cookies, 24 are also round. Imagine you take out a pile of round cookies. How many of them are salty cookies? ___ out of ____.
3. The principal of a school announced and explained a new school rule to all the students gathering together on the playground. Then the principal said: "Those who understand what I mean, please put up your hands." 70 out of every 100 students understood. Of these 70 who understood, 63 put up their hands. Of the remaining 30 who didn't understand, 9 put up their hands. Imagine a group of students who put up their hands. How many of them understood the principal? ___ out of ___.
4. 20 out of every 100 children in a school have bad teeth. Of these 20 children who have bad teeth, 10 love to eat sweet food. Of the remaining 80 children who don't have bad teeth, 24 also like to eat sweet food. Here is a group of children from this school who love to eat sweet food. How many of them may have bad teeth? ___ out of ___.
5. To protect their children's eyes, mothers always urge children not to watch too much TV. Suppose you want to test this belief and get the following information: 30 out of every 100 children become near-sighted. Of these 30 near-sighted children, 21 of them watch too much TV. Of those 70 children with normal sight, 28 of them watch too much TV. Suppose you meet a group of children who watch too much TV, how many of them may become near-sighted? ___ out of ___.
6. In Dongdong's town, 10 out of every 100 children are overweight. Of the 10 overweight children, 3 of them have overweight mothers. Of the remaining 90 children who have normal weight, 18 of them still have overweight mothers. Suppose you meet a group of overweight mothers in the town. How many of them have overweight children? ___ out of ___.
7. A group of children are playing games with cards. Those who get a card with a picture of a cat on the inner side win a piece of candy. 30 of every 100 cards have a cat picture on one side. Of the 30 cards with a cat picture, 12 of them are red on the other side. Of the remaining 70 cards that have no cat pictures, 35 of them are still red on the other side. Imagine Dingding takes out a group of red cards. How many of them have a cat picture on the other side? ___ out of ___.

8. In a cold winter in a town, 40 out of every 100 people hurt their hands by the cold. Of the 40 people who hurt their hands, 36 wear gloves in the open air. Of the remaining 60 people with normal hands, 30 also wear gloves. Suppose you meet a group of people who wear gloves in the town. How many of them hurt their hands? ___ out of ___.
9. In a hospital, 60 out of every 100 patients get a cold. Of the 60 patients who get a cold, 42 have a headache. Of the remaining 40 patients with other diseases, 12 also have a headache. Suppose you meet a group of patients who have a headache in a hospital. How many of them get a cold? ___ out of ___.
10. On a campus, 90 out of every 100 young people you meet are college students of this university. Of the 90 college students, 45 wear glasses. Of the remaining 10 young people that are not students of the university, 3 also wear glasses. Suppose you meet a group of young people who wear glasses on the campus. How many of them are students at this university? ___ out of ___.

References

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, *54*, 695–701.

Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillmann, K.-J., & Weiß, Hrsg. (2001). *PISA 2000 Zusammenfassung zentraler Befunde [Summary of key results]*. Berlin: Max Planck Institute for Human Development.

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

Brase, G. B. (2002). Ecological and evolutionary validity: Commment on Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni's (1999) mental model theory of extensional reasoning. *Psychological Review*, *109*, 722–728.

Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, *127*, 3–21.

Butterworth, B. (1999). *What counts: How every brain is hardwired for math*. New York: Free Press.

Butterworth, B. (2001). Statistics: What seems natural? *Science*, *292*, 853–854.

Carpenter, T., Coburn, T. G., Reys, R. E., & Wilson, J. W. (1978). *Results from the first mathematics assessment of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.

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.

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

Dawes, R. M. (1986). Representative thinking in clinical judgment. *Clinical Psychology Review*, *6*, 425–441.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974.

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

Evans, J. S., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, *77*, 197–213.

Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology*, *34*, 1340–1357.

Feynman, R. P. (1967). *The character of physical law*. Cambridge, MA: MIT Press.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon and Schuster (UK version: Reckoning with risk: Learning to live with uncertainty. London: Penguin).

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

Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren and Mellers and McGraw. *Psychological Review*, *106*, 425–430.

Gigerenzer, G., & Richter, H. R. (1990). Context effects and their interaction with development: Area judgments. *Cognitive Development*, *5*, 235–264.

Gigerenzer, G., Switjink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge, UK: Cambridge University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases. The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press.

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition*, *78*, 247–276.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*, 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., Lindsay, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*, 2261–2262.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*, 62–88.

Juslin, P., Winman, A., & Olssen, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review*, *107*, 384–396.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49–81). New York: Cambridge University Press, 49–81.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Kleiter, G. D. (1994). Natural sampling: Rationality without base rates. In G. H. Fischer, & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375–388). New York: Springer, 375–388.

Koehler, J. J. (1996a). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*, 1–53.

Koehler, J. J. (1996b). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review*, *67*, 859–886.

Krauss, S., & Atmaca, S. (2004). Wie man Schülern Einsicht in schwierige stochastische Probleme vermitteln kann. Eine Fallstudie über das "Drei-Türen-Problem" [How to give pupils insight into difficult stochastic problems. A case study on the "Three Door Problem"]. *Unterrichtswissenschaft*, *32*, 38–57.

Krauss, S., Martignon, L., & Hoffrage, U. (1999). Simplifying Bayesian inference: The general case. In L. Magnani, N. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 165–179). New York: Plenum Press, 165–179.

Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General*, *132*, 3–22.

Labarge, A. S., McCaffrey, R. J., & Brown, T. A. (2003). Neuropsychologists' ability to determine the predictive value of diagnostic tests. *Clinical Neuropsychology*, *18*, 165–175.

Laplace, P. S. (1951). *A philosophical essay on probabilities*. New York: Dover (Original work published 1814).

Lopes, L. (1991). The rhetoric of irrationality. *Theory and Psychology*, *1*, 65–82.

Lücking, A. (2004). *The development of Bayesian reasoning in children*. Unpublished diploma thesis, Free University, Berlin.

Martignon, L., Vitouch, O., Takezawa, M., & Forster, M. R. (2003). Naive and yet enlightened: From natural frequencies to fast and frugal decision trees. In D. Hardman, & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 189–211). Chichester, UK: Wiley, 189–211.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. New York: Norton (Original work published 1951).

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: Theory and evidence. *Annual Review of Ecological Systems*, *17*, 371–390.

Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development*, *72*, 103–122.

Schlottmann, A., & Anderson, N. H. (1994). Children's judgments of expected value. *Developmental Psychology*, *30*, 56–66.

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical applications*. Mahwah, NJ: Erlbaum.

Shrager, J., & Siegler, R. S. (1998). A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*, 405–410.

Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Science*, *3*, 430–435.

Siegler, R. S. (2000). The rebirth of children's learning. *Child Development*, *71*, 26–35.

Silver, E. A. (1986). Using conceptual and procedural knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 181–198). Hillsdale, NJ: Lawrence Erlbaum, 181–198.

Starkey, P., & Cooper, R. G. J. (1980). Perception of numbers by human infants. *Science*, *210*, 1033–1035.

Stern, E., Rode, C., Ge, F., Zhu, L. (2001). More than just numbers: Active diagrammatic competencies in Chinese and German secondary school students. *Poster session presented at the Biennial Meeting of the Society for Research in Child Development*, Minneapolis, MI.

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

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–750.