TRANSPARENT URNS AND COLORED TINKER-CUBES FOR NATURAL STOCHASTICS IN PRIMARY SCHOOL

Laura Martignon, Kathryn Laskey and Elke Kurz-Milcke

Institute for Mathematics, Pädagogische Hochschule Ludwigsburg

and the Krasnow Institute, George Mason University

We join the camp of mathematics educators who claim that children should receive early training in stochastic thinking, although their training – at such stage – may only be based on a heuristic understanding of stochastic phenomena. The term "heuristic" is to be taken in the sense propagated by scientists like Einstein and Polya, which is not as "a rule of thumb" but as a correct yet partial approximation of the normative approach. We present a review of empirical results supporting our claim and propose to guide children to construct stochastic situations enactively.

EFFECT OF REPRESENTATION FORMATS IN PROBABILISTIC REASONING TASKS

The theoretical framework of this paper is that provided by Hasher & Zack's work on humans' automatic recording of frequencies in the environment combined with Barsalou's mental concept simulators (1999) and the theory of heuristics for inference proposed by Gigerenzer, Todd and the ABC Group (1999). Barsalou's simulators implement a basic conceptual system that represents types, supports categorization, and produces categorical inferences. Productivity results from integrating simulators combinatorially and recursively to produce complex simulations. In this framework mathematics education of young students consists in the education of an inner representation space where mental simulators of mathematical processes are implemented (e.g., "imagine drawing from an urn"). Mathematical intuitions in general and probabilistic intuitions in particular are thus conceptually replaced by heuristics for inference combined with mental simulations that are part of an adaptive toolbox (Gigerenzer et al. 1999). For such mental simulators *natural* frequencies (see below for a characterization), are more *ecologically rational* than percentages and probabilities (1995). We are beginning to find confirmations from cognitive neuroscience of this thesis. In fact, we have collected evidence suggesting that the regions of the brain which are active when we perform probabilistic inferences by means of natural frequencies differ from those that are active when we solve probabilistic problems with percentages or probabilities. This is true even for expressions such "1 out of 4" as compared with expressions such as "0.25" or "1/4".

Initial experimental results substantiating the above hypotheses support an emphasis in schools on *natural* representation formats for probabilistic information. Here, the term "natural" means arising either directly from enactively constructing subcategories of a population by partitioning it sequentially in nested subsets and determining the proportions of the subcategories thus formed, or mentally simulating these same processes. Our emphasis is not meant to replace instruction in percentages or measure theoretic probability. On the contrary, we view early school interventions as a means to prepare young children for later instruction in working in the formal mathematics of probability. Enactive learning approaches, we claim, can teach young children to reason with proportions of counted items, thus making use of natural representation formats to develop intuitions. Probabilistic inference in secondary school can then make use of this previously acquired substrate by anchoring probabilistic reasoning in "translations" of probabilities into natural frequencies. The advantages of such "translations" have been empirically tested in interventions in secondary school (DFG Project in BIQUA, Ma-1544/1-4 and Bi-384/4-3).

Cognitive Processes in Probabilistic Reasoning Tasks

As awareness grows of the importance of uncertainty in everyday life and in public affairs, concern also grows about the competence of the citizenry to process uncertainty in a sound and effective manner. An alarmingly large proportion of the public cannot make effective use of probabilistic information. The German newspaper "Süddeutsche" (Süddeutsche Zeitung Magazin, 31.12.1998) asked 1000 Germans, what they think is the meaning of 40%: a quarter, 4 out of 10, or every 40th. Only 54% knew the correct answer, which is "4 out of 10". A burgeoning literature has documented disparities between the results of unaided judgment and the prescriptions of the probability calculus (KAHNEMAN, SLOVIC AND TVERSKY, 1982). Summarizing the literature on human performance on probabilistic reasoning tasks, Gould (1992) commented: "*Tversky and Kahneman argue, correctly, I think, that our minds are not built (for whatever reason) to work by the rules of probability.*" Yet, recent re-examinations of the literature on human performance on tasks involving uncertainty have concluded that to a large extent, the negative results can be explained by discrepancies between the environment and tasks on which present-day humans perform so poorly, and those faced by our ancient forebears (e.g., GIGERENZER, et al., 1999). The question is then: *Would it be possible to improve the public's skill at probabilistic reasoning by matching pedagogical strategies adaptively to cognitive processes during early phases of education, thus providing anchoring mechanisms and "translation" heuristics for the phase when more formal representations are taught?* We build on a base of existing results on the cognitive mechanisms underlying probabilistic reasoning. Our research was originally motivated by an important type of probabilistic reasoning task known as "Bayesian reasoning." A prototypical Bayesian reasoning task involves using evidence about an uncertain proposition to revise our assessment of the likelihood of a related proposition. The following example is drawn from a recent article by Zhu and Gigerenzer (2006) that examined children's ability to perform Bayesian reasoning. The context was a small village in which "red nose" was a "symptom" of "telling lies". The task required the children to relate the proposition "having a red nose" to the proposition "telling lies." Specifically, they were given information about the probability of "red nose" conditioned on "liar" and that of "red nose" conditioned on "non-liar," as well as the incidence of liars in the village. They were then asked to establish the chance that someone with a red nose tells lies. The performance of children (fourth graders) improved substantially when the probabilistic setting was replaced by a setting in which the cover story reported the natural frequencies involved, i.e., in terms of a sequential partitioning of nested sets and their proportions. Humans, children and adults, appear to be adapted to a "natural" sequential partitioning for categorization. The term "natural" means that these proportions (i.e., relative frequencies) are perceived as being obtained by the mental simulation of counting. ATMACA AND MARTIGNON (2004) conjectured that different neural circuits are involved in the natural frequency and probability versions of the Bayesian task. They reported experimental results that support their conjecture. Subjects were given tasks by slide projector, and solved them mentally with no writing allowed. Information was collected on correctness of solutions and the time to solution. The experiment made use of a response mode called *result verification* or *result disparity* (KIEFER and DEHAENE, 1997): Subjects are presented with a proposed solution and asked to judge as quickly as possible whether it is correct or incorrect. ATMACA AND MARTIGNON found that subjects needed significantly longer times and produced significantly fewer correct answers, for the tasks given in probability format versus those given in the natural frequency format. In one experiment 110 participants were exposed to typical Bayesian tasks with "three branches" and tasks with "four branches" as represented below:

Three branches: 10 out of 1000 children have German measles . Out of the 10 children who have German measles, all 10 have a red rash. *Of the 990 children without German measles, 9 also have a red rash. How many of the children with a red rash have the German measles?*

Four branches: *10 out of 1000 car drivers meet with an accident at night. Out of the 10 car drivers who meet with an accident at night, 8 are intoxicated. Out of the 990 car drivers who do not meet with an accident at night, 40 also are intoxicated. How many of the car drivers who are intoxicated actually meet with an accident at night?*

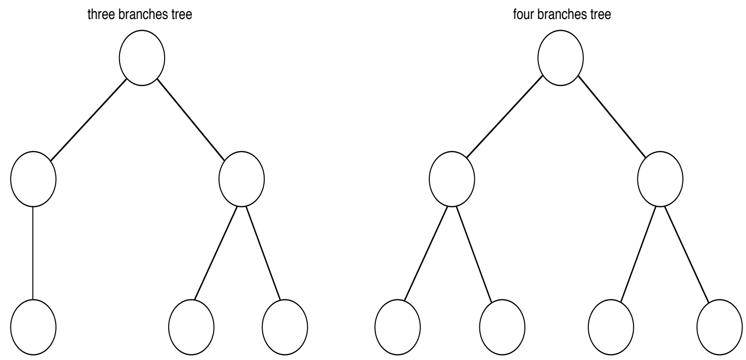


Figure 1: Bayesian Task

Results of the experiment are summarized in the following Figure:

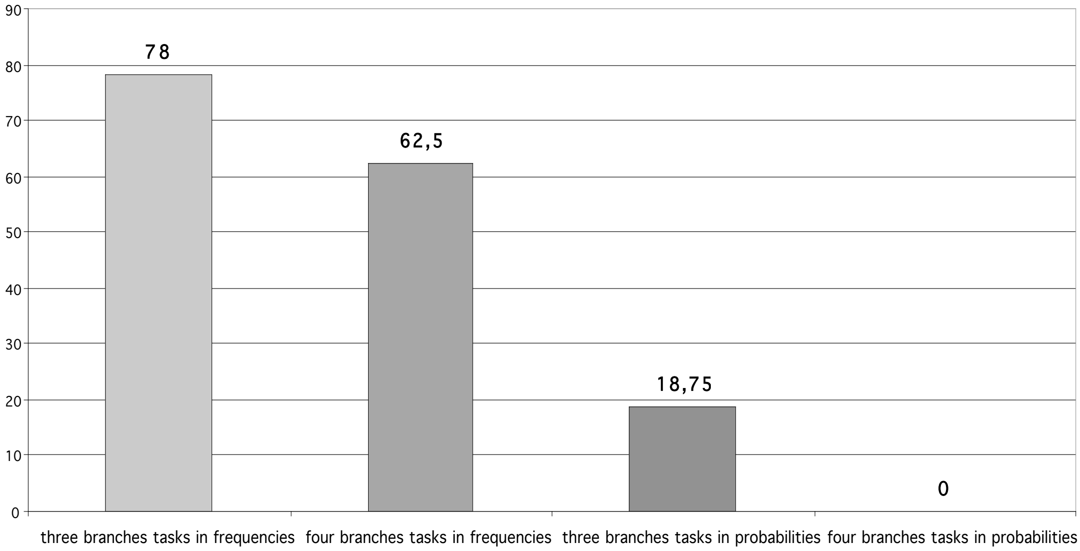


Figure 2: Percentage of Subects Solving Task Correctly

Perception of frequencies of occurrences, this result suggests, could be a mechanism, or at least part of a more complex mechanism, that enables fast and effective decisions in uncertain situations, because "*…natural selection (…) gives rise to practical cognitive mechanisms that can solve (…) real world problems...*" (FIDDICK & BARRETT, 2001, S. 4). DEHAENE (1997) wrote in a similar context: "*Evolution has been able to conceive such complex strategies for food gathering, storing, and predation, that it should not be astonishing that an operation as simple as the comparison of two quantities is available to so many species).*" (DEHAENE 1997, p. 27)

The Automatic Processing of Frequencies

The automatic perception of frequencies of occurrences was described by HASHER AND ZACKS in the 70s: "Operations that drain minimal energy from our limited-capacity attentional mechanism are called automatic; their occurrence does not interfere with other ongoing cognitive activity. They occur without intention and do not benefit from practice". Certain automatic processes, we propose, are ones for which humans are genetically "prepared". These processes encode the fundamental aspects of the flow of information, namely, spatial, temporal, and frequency-of-occurrence information." (HASHER AND ZACKS, 1979, p. 356). Humans are known to be sensitive to frequencies even when they do not pay attention to them. In experiments, participants performed well when remembering frequencies of events, even when they had no reason to expect a memory test at all (ZACKS, HASHER AND SANFT, 1982), and there seems to be evidence that they did not count the events (COREN AND PORAC, 1977). As reported by HASHER AND ZACKS (1984), several experiments have shown that neither activated intention, nor training, nor feedback, nor individual differences such as intelligence, knowledge or motivation, nor age, nor reductions in cognitive capacity such as depression or multiple task demand, have an influence on the processing of frequencies of occurrence information. That is to say, there are strong hints that the human brain adapted to the processing of frequencies during evolution. Furthermore, several experiments have provided evidence that "(…) various animal species including rats, pigeons, raccoons, dolphins, parrots, monkeys and chimpanzees can discriminate the numerosity of various sets, including visual objects presented simultaneously or sequentially and auditory sequences of sounds" (DEHAENE ET AL., 1998, p. 357). According to models of animal counting presented by MECK AND CHURCH (1983), numbers are represented internally by the continuous states of an analogue accumulator. For each counted item, a more-or-less fixed quantity is added to the accumulator. The final state of the accumulator therefore correlates well with numerosity, although it may not be a completely precise representation of it. (DEHAENE, 1992) This model explains the observation that animals are very good in handling small quantities, while performance degrades with the increase of magnitude. During this counting process, "the current content of the accumulator is used as a representative of the numerosity of the set so far counted in the decision processes that involve comparing a current count to a remembered count." (GALLISTEL AND GELMAN, 1992, p. 52) That is to say, the current content of the accumulator represents the magnitude of the current experienced numerosity, whereas previously read out magnitudes are represented in long-term memory. This enables comparison of two number quantities, which is essential for frequency processing. The comparison of number processing abilities in animals and human infants leads to the conclusion "that animal number processing reflects the operation of a dedicated, biologically determined neural system that humans also share and which is fundamental to the uniquely human ability to develop higher-level arithmetic." (DEHAENE ET AL., 1998, p. 358). Several studies report abilities of frequency perception in kindergartners and elementary school children in the range of grades 1 to 6 (HASHER AND ZACKS, 1979; HASHER AND CHROMIAK, 1977). And the findings of abilities in numerosity discrimination in infants and even newborns (ANTELL AND KEATING, 1983) "may indicate that some capacity for encoding frequency is present from birth." (HASHER AND ZACKS, 1984, p. 1378).

From Absolute Numerical Quantities to Natural Frequencies

Animals and humans could not survive if they had only developed a sense for absolute frequencies without a sense for *proportions of numerical quantities* for inference. "Are all red mushrooms poisonous, or only some of them? How valid is red colour as an indicator of poison danger in the case of mushrooms?" Whereas non-precise estimates may have been sufficient for survival in ancient rural societies, answering this type of question by means of well calibrated inferences is vital in modern human communities. Successful citizenry requires this type of competency and, it is our conviction, elementary school should provide tools for successful quantified inferences. In order to establish whether one cue is a better predictor than another (e.g., whether red colour is a better predictor than white dots for poisonous mushrooms) we need a well tuned mental mechanism that compares proportions. Little is known so far, in the realm of cognitive neuroscience, as to which brain processes are involved in proportion estimation and proportion comparison. The more rudimentary instruments for approximate inference that humans share with animals must be based on some sort of non-precise proportion comparison (GIGERENZER, ET AL., 1999). But when do infants in modern societies begin to quantify their categorization and when can they be trained in quantified proportional thinking? Although fractions are the mathematical tool for describing proportions, we envisage an early preparation of children's use of numerical proportions without "normalizations" before they are confronted with fractions. The results by Piaget and Inhelder on the understanding of such proportions in children motivated two generations of researchers in developmental and in pedagogical psychology related to mathematics education. We cite here one direction in particular, which has been fundamental to our work (Koerber, 2003). In a series of well designed experiments, STERN, KOERBER and colleagues (STERN ET AL. 2002) demonstrated that third-graders can *learn to abandon* the so-called *additive misconception*, in which children respond with "9" instead of "12" to "3 : 6 = 6 : ?". In these experiments, children were asked to compare mixtures of lemon and orange with respect to their intensity of taste. The training involved using a balance beam or graphs to represent juice mixtures, and moving the pivot to represent changes in proportions. At the end of a short training (2 days at most) children showed improvement in proportional thinking, The results of STERN and her school thus provide evidence of third-graders' aptitude to learn proportional thinking when provided with adequate instruction.

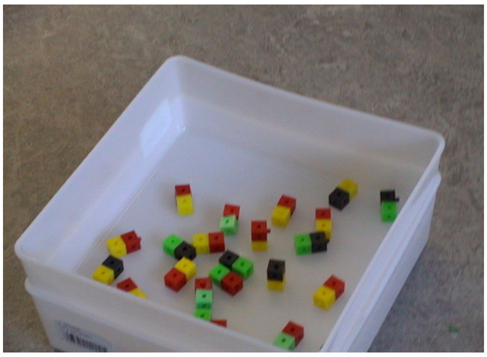
Cognitively Natural Representations and Task Performance

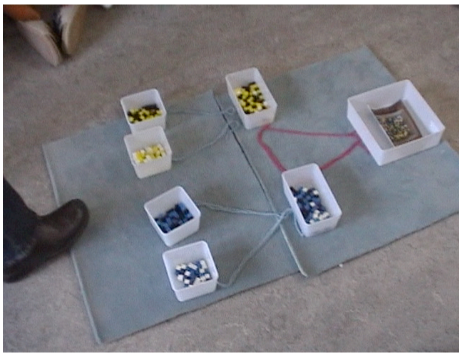
Stern's results can be combined with results of Gigerenzer and his school at the interface between pedagogy and cognitive science in the search for pedagogical approaches that tap into cognitively natural representations. The *natural frequency* representation for Bayesian reasoning tasks is based on information that can be gained by "naturally"counting events in an environment, and therefore taps into very basic human information processing capacities. "*Natural sampling is the way humans have encountered statistical information during most of their history. Collecting data in this way results in natural frequencies*." (HOFFRAGE, GIGERENZER, KRAUSS & MARTIGNON, 2004) The term "natural", as has been pointed out, signifies that these frequencies have not been normalized with respect to base rates. Probabilities and percentages can be derived from natural frequencies by normalizing natural frequencies into the interval [0,1] or [0, 100], respectively; however, this transformed representation results in loss of information about base rates. Consider the following examples from HOFFRAGE, GIGERENZER, KRAUSS & MARTIGNON, 2004:

> Natural frequencies: *Out of each 100 patients, 4 are infected. Out of 4 infected patients, 3 will test positive. Out of 96 uninfected patients, 12 will also test positive.*

Normalized frequencies: *Out of each 100 patients, 4 are infected. Out of 100 infected patients, 75 will test positive. Out of 100 uninfected patients, 12.5 will also test positive.*

Because normalized frequencies filter out base rate information, they make the Bayesian task of inferring a posterior probability from evidence more difficult. The inference that most positives are false positives can be read directly from the natural frequency representation, while it must be obtained via a non-trivial calculation from the normalized frequency representation. Training young children enactively with natural proportions, we claim, enables them to make use of simple heuristics for dealing with probabilities. We have worked with fourth graders, preparing them to solve one of the mathematical test items of PISA 2003 namely: *Consider two boxes A and B. Box A contains three marbles, of which one is white and two are black. Box B contains 7 marbles, of which two are white and five are black. You have to draw a marble form one of the boxes with your eyes covered. From which box should you draw if you want a white marble?* Only 27% of the German school students were able to justify that one should choose Box A. Mathematically correct statements regarding why and when larger proportions in samples correspond to larger chances in the populations require serious amounts of conceptual work. In early grades, the consensus is that one should focus on intuition and competency rather than on formal mathematics. In other words, we should provide students with: (1) Basic stochastic modelling skills with natural representation formats and (2) Simple heuristics for operating with these formats. In this spirit, MARTIGNON AND KURZ-MILCKE (2005) and KURZ-MILKE AND MARTIGNON (2005) have designed a program that develops and encourages the natural frequency representation through the use of enactive learning. In playful yet structured activities, children use coloured plastic cubes called tinker-cubes to represent individuals that make up a population. Different colours represent different attributes (e.g., red cubes for girls; blue for boys). The cubes can be attached one to another, allowing representation and multi-attribute encoding (e.g., a red cube attached to a yellow cube for a girl with glasses; a blue cube attached to a green cube for a boy without glasses, a red cube attached to a green cube for a girl without glasses, and so on). Children collect the tinker-cubes into plastic urns that represent populations. In this way, they gain concrete visual and tactile experience with individuals with multiple attribute combinations and how they can be grouped into categories and subcategories. Recent exploratory studies of fourth-grade children indicate that children are both enthusiastic and successful when constructing these representations of categories and sub-categories in nested sets, especially when the populations are personally meaningful (e.g., “our class”). They can easily “construct” answers to questions like “how many of the children wearing glasses are boys?”





2. Constructing our class

3. Enactive

Bayesian

Reasoning

4th – graders

(Kurz-Milcke&Martignon, 2005)

Figure 2

Figure 3

In another activity, children enact a model of proportional reasoning by constructing so called similar urns to represent equivalent proportions (e.g., an urn containing 2 red and 5 blue tinker-cubes – denoted by U (2:5) - is similar to an urn containing 4 red and 10 blue tinker-cubes). By this activity the children learn basic urn arithmetic. For instance, fourth graders in three classes of a school in Stuttgart successfully learned to solve the two boxes task described above, where U(1:2) is compared with U(2:5) by first constructing an urn U(2:4) similar to U(1:2) and then easily comparing U(1:2) and U(2:5). These tasks contain first elements of elementary probabilistic reasoning in general but also of Bayesian reasoning at a heuristic level. They represent a preparation for understanding both of fractions and – at a later stage – of probabilities. Empirical longitudinal studies have now been designed to confirm the hypothesis that mastery of these tasks in the younger grades should support better performance on stochastics questions in the later grades.

REFERENCES

Atmaca, S. & Martignon, L. (2004) Hat sich unser Gehirn an die Wahrnehmung und an die Verarbeitung von Häufigkeiten adaptiert? *Beiträge zur 46 Tagung experimentell arbeitender Psychologen* Gießen, 2004, 134-138

Antell, S.E. and Keating, D.P. (1983). Perception of Numerical Invariance in Neonates. *Child Development, 54*, pp. 695-701.

Barsalou, L.(1999) Perceptual symbol systems. Behavioral and brain science. 22. 577-660

Church, R.M. and Meck, W.H. (1984). In H.L. Roitblat, T.G. Bever, and H.S. Terrace, Eds.: *Animal Cognition*. Erlbaum.

Corbetta, M, Miezin, F.M., Schulman, G.L., and Petersen, S.E. (1993). A PET Study of Visuospatial Attention. *Journal of Neuroscience, 13* (3), pp. 1202-1226.

Coren, S., and Porac, C. (1977). Fifty Centuries of Right-Handedness: The Historical Record. *Science, 198*, 631-632.

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.

DFG Projekt „Entscheidungsfindung unter Unsicherheit als fächerübergreifende Kompetenz – Alltagsorientierter Stochastikunterricht am Gymnasium (Ma

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, pp. 1-42.

Dehaene, S. (1997). *The Number Sense*. Oxford University Press.

Dehaene, S., Dehaene-Lambertz, and Cohen, L. (1998). Abstract Representations of Numbers in the Animal and Human Brain. *TINS, 21*, 8, pp. 355-361.

Fiddick, L. and Barrett, H.C. (2001). Evolution of Cognition: An Adaptationist Perspective. In: N.J. Smelser and P.B. Baltes (Eds.): *International encyclopedia of the social and behavioral sciences*. Vol. 7. Amsterdam: Elsevier, 2001, pp. 4996-5000.

Gallistel, C.R. and Gelman, R. (1992). Preverbal and Verbal Counting and Computation. *Cognition, 44*, pp. 43-74.

Gigerenzer, G. (2004). Calculated risks: What Numbers Really Tell Us About Our Lives. Simon and Schuster, New York.

Gigerenzer, P.M. Todd & the ABC Research Group (1999) Simple Heuristic that make us smart. New York: Oxford University Press.

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

Goldstein, D., Hasher, L., and Stein, D.K. (1983). The Processing of Occurence Rate and Item Information by Children of Different Ages and Abilities. *American Journal of Psychology, 96*, pp. 229-241.

Hasher, L. and Chromiak, W. (1977). The Processing of Frequency Information: an Automatic Mechanism? *Journal of Verbal Learning and Verbal Behavior, 16*, pp. 173-184.

Hasher, L. and Zacks, R.T. (1979): Automatic and effortful processes in memory. *Journal of Experimental Psychologie: General, 108*, 356-388.

Hasher, L. and Zacks, R.T. (1984). Automatic Processing of Fundamental Information - The Case of Frequency of Occurrence. *American Psychologist, 39*, 12, pp. 1372-1388.

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.

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

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological review, 80,* 237-251.

Kiefer, M. and Dehaene, S. (1997). The Time Course of Parietal Activation in Single-digit Multiplication: Evidence from Event-related Potentials. *Mathematical Cognition, 3*, 1, pp. 1-30.

Koerber, S. (2003) Visualisierung als Werkzeug im Mathematischen Unterricht: Der Einfluß externer Repräsentationen auf Proportionales Denken im Grundschulalter. Dr. Kovac Verlag.

Kurz-Milcke, E., & Martignon, L. (2005) Lebendige Urnen und Erlebnissreiche Bäume in der Grundschule. „Berichte des Arbeitskreises Stochastik in der Schule“ Nr. 3 Franzbecker: Hildesheim

Martignon, L. & Kurz-Milcke, E. (2005) Haufen bunter Steckwürfel für eine natürliche, gendersensitive Stochastik in der Grundschule Berichte des Arbeitskreises Stochastik in der Schule Nr. 3. Franzbecker: Hildesheim

Martignon, L.& Wassner, Ch. (2005) Schulung frühen stochastischen Denkens von Kindern Zeitschrift für Erziehungswissenschaft, 8. Jahrg., Heft 2, S.2002-222

Meck, W.H. and Church, R.M. (1983). A Mode Control Model of Counting and Timing Processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*, 3, pp. 320-334.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., Fazio, F. (1999). Distinct Brain Loci in Deductive Versus Probabilistic Reasoning. *Neuropsychologica, 36*, 4, pp. 369-376.

Stern, E., Hardy, I. & Koerber, S. (2002) Die Nutzung graphisch-visueller Repräsentationsformen im Sachunterricht.In K. Spreckelsen, A. Hartinger, & K. Möller (Eds.) Ansätze und Methoden empririscher Forschung zum Sachunterricht (pp. 119-131). Bad Heilbrunn: Klinkhardt.

Wassner, C., Martignon, L., Biehler, R. (2004). Bayesianisches Denken in der Schule. In: Unterrichtswissenschaft. Zeitschrift für Lernforschung, 32. Jahrgang, Heft 1 (pp. 58-96). Juventa Verlag, Weinheim.

Zacks. R.T., Hasher, L., and Sanft, H. (1982). Automatic Encoding of Event Frequencies - Further Findings. *Journal of Experimental Psychology - Human Learning and Memory, 8* (2), pp. 106-116.

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