Probabilistic Judgment by a Coarser Scale: Behavioral and ERP Evidence

Yanlong Sun (Yanlong.Sun@uth.tmc.edu)[1]
Hongbin Wang (Hongbin.Wang@uth.tmc.edu)[1]
Yingrui Yang (yangyri@rpi.edu)[2]
Jiajie Zhang (Jiajie.Zhang@uth.tmc.edu)[1]
Jack W. Smith (Jack.W.Smith@uth.tmc.edu)[1]

[1] The University of Texas Health Science Center at Houston, Houston, TX 77030 USA
[2] Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY 12180 USA

Abstract

We hypothesize that statistically unsophisticated people perceive event likelihood with a coarser scale with a limited number of categories, before they report exact numerical probability values. Refinement of the scale beyond a certain level would therefore not improve overall judgmental accuracy and consistency but just impose a heavier burden on their limited computational capacity. An experiment of probabilistic judgment was conducted to test this hypothesis. Results from both behavioral data and event-related potentials in EEG recordings supported our hypothesis.

Introduction

Assessing the likelihood of uncertain events is an essential aspect of human reasoning and decision-making. In the absence of adequate formal models for computing the probabilities, people often rely on intuitions and heuristics to assess uncertainty. The question of how lay people and experts evaluate the probabilities of uncertain events has attracted enormous research interest. (For a historical review, see, e.g., Goldstein & Hogarth, 1996). Proponents of the "heuristics and biases" program argued that intuitive probabilistic judgment is often systematically biased and error-prone (Kahneman, Slovic, & Teversky, 1982). Various violations of normative models, including overconfidence, base-rate neglect, and the conjunction fallacy, have been attributed to applications of a small number of distinctive judgmental heuristics. However, others argued that the human mind has evolved to deal with the structure of the social and physical environment rather than to solve abstract probability problems (Chase, Hertwig, & Gigerenzer, 1998; Gigerenzer et al., 1999; Wang, Johnson, & Zhang, in press). When external information is presented effectively, people can be good intuitive statisticians (e.g., Cosmides & Tobby, 1996). For instance, Gigerenzer and colleagues found that when problems were stated in terms of frequencies instead of probabilities, the stable errors of judgments disappeared (Gigerenzer, 1991; Sedlmeier & Gigerenzer, 2001). For a recent collection of different perspectives on interpreting human intuitive probabilistic judgment, see Gilovich, Griffin, & Kahneman, 2002.

The theoretical claim that the human mind is not adapted to process probabilities has been a magnet for controversy (e.g., Gigerenzer, 1994; Gigerenzer & Hoffrage, 1995; Kahneman & Tversky, 1996). In spite of the centrality of the question, the internal representations and functional neural foundations underlying human probabilistic judgment are poorly understood. Until recently, most research has relied on observations and interpretations of behavioral experiments. In this paper we report a study that investigates the internal representations of human intuitive probabilistic assessment, from both behavioral and neurological perspectives.

Uncertainty Assessment by Approximation

A common scheme in psychological experiments of probabilistic judgment is to ask participants to report probabilities in numerical values, such as the chance of breast cancer in percentage given a positive test result (e.g., Sedlmeier & Gigerenzer, 2001). One question with this scheme, however, is that participants' response of numerical probability values may not genuinely reflect the true internal representation of their likelihood assessment. The following example illustrates this point. When asked to give a numerical estimate of the probability of a certain event provided with probabilities of other events, people often give incoherent answers (e.g., Osherson et al., 2001):

(a) Prob (Clinton is re-elected to the Senate in 2006) = .75
(b) Prob (Giuliani runs for the Senate in 2006) = .5

Participants' response:
(c) Prob (Clinton is re-elected to the Senate in 2006 and Giuliani runs for the Senate in 2006) = .1

Apparently, the numerical estimate of .1 is incoherent with the other two given probabilities, and the correct answer should be at least .25. However, it is possible that participants did not distinguish small increments on the continuous scale of probabilities. The estimate of .1 probably is only an approximation by the idea that the chance for Clinton may drop dramatically if Giuliani joins in the competition, rather than the result of exact calculations. In other words, the incoherence is likely to be produced by the application of a coarser scale, rather than by some systematically biased heuristic.

We propose two hypotheses on the internal representations when people intuitively assess the event likelihood without exact calculations. First, the task of judgment of probabilities can be partitioned into two separate phases: the internal representation and the response (see Figure 1). The internal representation of event likelihood is the result of an approximate estimation of presented information (cues) on a coarser scale (internal scale). Only when subjects need to report probability values (e.g., in a typical psychological experiment), the internal scale is projected onto a finer continuous scale (response scale). Second, the internal scale has only a limited number of categories that represent different magnitudes of the perceived event likelihood. Together, these two hypotheses allow us to distinguish two fundamentally distinctive types of internal representations underlying human probabilistic judgment. We illustrate each of them in detail next.

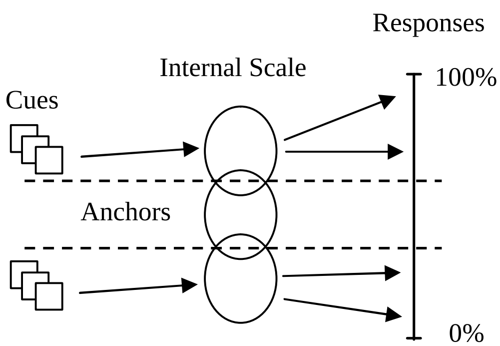

Figure 1. Internal representation on a coarser scale

Internalization by Anchoring and Adjustment

The hypothesis that people's perception of the event likelihood may not be infinitely refined is based on the notion that human beings have limited computational capacity. Miller (1956) suggested that the number of levels of any variable that can be internalized is not only finite but also small. Miller's finding has been widely cited in psychometric research on whether there is an optimal number of response alternatives in designing a scale. Many researchers believe that a further refinement of scales is meaningless if it is beyond human information processing capacity (for a review, see Cox, 1980).

In assessing the event likelihood, the number of categories on the internal scale is determined by the way these categories are formed. Previous studies show that people often use the "anchoring and adjustment" heuristic in judgments of belief and value (Tversky & Kahneman, 1974). Tversky and Kahneman presented anchoring as a process in which "people make estimates by starting from an initial value that is adjusted to yield a final answer" (1974, p. 1128). Chapman & Johnson (2002) provided a review on recent development in understanding the mechanisms of anchoring. One important implication of anchoring is that the number of categories on the internal scale is limited, rather than infinitely refined, since there are only a limited number of anchors that can be processed at the same time (see Figure 1). Note that it is not to suggest that people can not distinguish two very alike events with subtle difference. However, such distinction can only be achieved based on a side-by-side comparison, as one of the two events serves as an anchor.

The anchors can be directly provided by external cues, predefined action thresholds, or by knowledge retrieved from memory (for example, availability heuristic, Tversky & Kahneman, 1974). In the example of incoherent estimate mentioned before, the anchors would be likely provided by the probabilities stated in the first two statements, .75 and .5. Thus, there would be only three categories of likelihood divided by these two numbers, more likely than .75, less likely than .5, and a category in the between. When asked to report a numerical probability estimate to the third statement, the category of the lowest possibility was projected onto a continuous scale As a result, a small number such as .1 was likely to be generated. It is very conceivable, however, that in a different occasion the same person would give an answer of .2 instead of .1 to the same question without changing his or her perception of the event.

In everyday life situations, it is often the case that one of a few options is chosen based on predefined action thresholds rather than on exact probability values. It was found that physicians often significantly overestimated the probability of disease given a positive test result. In one study, 95 out of 100 physicians estimated the probability of breast cancer between 70% and 80% given a positive mammography while the correct answer is merely 7.8% (Eddy, 1982). Nevertheless, such large deviations do not necessarily mean that they are poor physicians. Probably to a physician, what matters most is a dichotomy whether a test result is positive. Consequently, a two-category internal representation (for example, "more likely" and "less likely") would be adopted until further diagnosis is conducted. When the physicians were asked to report an exact numerical value, they simply just obtained a number that would represent the category of the highest likelihood on the internal scale. Thus, a large number was likely to be reported.

Coarser Scales versus Finer Scales

The above argument actually suggests the need to partition the errors of intuitive probabilistic judgment. By distinguishing the internal representation from the numerical responses, we in effect partition the errors into two sources: systematic errors when the external information is internalized onto a coarser scale, such as overconfidence, availability bias, and conjunction fallacy (e.g., Kahneman, Slovic, & Teversky, 1982), and "random errors" when the internal representation are projected onto the continuous scale of numerical values. From Figure 1, it can be seen that a large portion of errors is elicited by the projection of a coarser scale onto a continuous scale. Thus, a direct comparison between human perception of event likelihood and criteria derived from a normative model very likely has exaggerated the irrationality of human intuitive judgment. In the previous example of incoherent estimate, incoherence would be greatly reduced if a coarser scale is used to evaluate participants' responses.

We reason that in assessing uncertainty, people prefer a coarser scale (e.g., fewer categories) to a finer scale (e.g., more categories) as the internal representation, since the former may function effectively and demand less computational load in a variety of cognitive tasks. People assess the event likelihood only to the extent that it is adequate to reach a conclusion or choose an action. For example, when a person needs to decide whether or not to bring an umbrella to work based on the chance of rain, a chance of 70% and a chance of 80% probably would not make much difference in such a decision. Sometimes a finer judgment can be obtained but such refinement might have little effect to improve the choice. For example, Kareev and colleagues suggested that the limited capacity of working memory could actually help the early detection of covariation (Kareev, 1995; Kareev, Lieberman & Lev, 1997). In Sun & Tweney (2002), researchers found that participants chose their actions based on a heuristic of using small samples, and their performance in the task was very close to that obtained by the optimal strategy.

Our hypothesis that there exists a coarse representation of uncertainty for intuitive probabilistic judgment is consistent with the large body of literature on mental presentations of quantity and numbers. Dehaene and colleagues (Dehaene, 1997; Dehaene et al, 1998) have suggested that there is a coarse and analog mental number line, which is the foundation of a "number sense" and shared by humans and animals. Recent studies using brain-imaging techniques have provided further support for the existence of such coarse scale representations. It has been found that the approximation and exact calculation tasks of large numbers (as compared to rote arithmetic operations) put heavy emphasis on the left and right parietal cortices, which may encode numbers in a non-verbal quantity format (e.g., Dehaene et al., 1999; Pesenti et al., 2000; Stanescu-Cosson et al., 2000).

To test our hypotheses, we conducted an experiment in which participants performed a task of probability estimation. The task was to estimate the winning probabilities of poker hands in a standard "draw poker" game when presented with the highest two cards out of the five cards on a hand. The reason we selected this task is that it provides an objective criterion for evaluating participants' estimates. Most importantly, this task offers a distribution of probabilities ranging from zero to close to 100% with extremely small increments. This feature allows us to look into the refinement of the internal scale when people need to make estimates intuitively. We compared two conditions in the experiment. The coarser-scale condition required probability estimation within an increment of 30% (e.g., less than 30%, 30% ~ 60%, and greater than 60%). The finer-scale condition required probability estimation within an increment of 10%. Both behavioral data and the event related potentials (ERP) in EEG recordings were collected. The experiment results supported our hypotheses. First, estimates by the finer scale showed significantly worse performance in terms of overall accuracy and consistency. Second, significant ERP difference was found over the parietal area between two conditions, indicating that different levels of effort were involved in the application of different internal scales.

Method

Participants Six graduate students in the Houston medical center area participated in the experiment. All participants were right-handed males. The averaged age was 27 years old. None of the participants reported having any in-depth knowledge of probabilistic theories. All participants reported having some experience of playing poker games but only at a novice level.

Procedure The stimuli in the experiment (two-card poker hands) consisted of 18 levels of winning probabilities, ranging from 5% to 90% with a step of approximately 5% (as in a one-deck and two-person game). With suit variations, there were 36 different hands used in the experiment. (Note that in this experiment, the suit variation does not change the winning probability since the highest hand is a one-pair of aces.) Participants were introduced with the rules of hand ranking before the experiment (e.g., a pair is higher than single cards; the Ace is higher than the King, and etc.) without being informed of the corresponding probability values.

Figure 2 shows the procedure of the experiment. Stimuli were displayed on a 17-inch CRT monitor. To reduce eye movements, the viewing angle of the displays on the screen was limited to 2%. After the fixation display, two poker cards were presented. Participants were instructed to form an estimate of winning probability in their mind immediately. After the second fixation, a number was displayed and the participants needed to compare their estimate to the displayed number as quickly as possible. A Microsoft PC mouse was used to collect the responses (the left button for "less than," and the right button for "greater than").

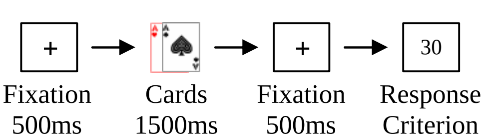

Figure 2. Experiment Procedure

There were two experimental conditions: the coarser-scale condition and the finer-scale condition. In the coarser-scale condition, participants were instructed to estimate probabilities in three categories (less than 30%, 30% ~ 60%, and more than 60%). The number displayed as the comparison criterion is either 30 or 60, randomly selected. In the finer-scale condition, participants were instructed to form the estimate accurate to the single digit. The comparison criterion was either 5% plus or 5% minus (randomly selected) the true winning probability. For example, a pair of aces has a winning probability of 90%. Then, participants needed to compare their estimates with either 85% or 95%, depending on which number was displayed. Each participant completed both conditions. The order of two conditions was balanced among participants. Each condition consisted of 72 trials in 2 trial list cycles, and each of the 36 poker hands was displayed once in each cycle. The order of trials was randomly shuffled. Response time was recorded as the latency after the onset of comparison criterion. During the experiment, ERPs were sampled at 250 Hz with a 128-electrode geodesic sensor net (GSN) reference to the vertex (Tucker, 1993).

Results

Behavioral Results Table 1 shows the comparison between two experimental conditions on participants' probabilistic estimates in three categories, accuracy, consistency, and response time. Accuracy was calculated as the percentage when participants made the comparison correctly. For example, if the true winning probability was 70 percent, and the number displayed as the comparison criterion was 60, the correct response should be "greater than." (In case of identical numbers, either response was counted as correct.) All participants showed higher levels of accuracy in the coarser-scale than in the finer-scale condition. On average, the accuracy was 76.4% in the coarser-scale condition, and 56.0% in the finer-scale condition, with a difference of 20.4% ($t(5) = 4.576$, $p < 0.01$, two-tailed). Furthermore, we compared participants' accuracy in two trial list cycles in each condition and found little improvement in accuracy over the trials. The low accuracy level in the finer-scale condition was not significantly different from random guess (comparing to the expected value of 50%, $t(5) = 1.836$, $p = .12$, two-tailed), indicating that participants were not able to distinguish winning probabilities of poker hands in the increment of 10%.

To evaluate participants' judgmental consistency, we calculated Pearson correlations in their responses over two identical sets of stimuli (36 poker hands in each set) in each condition. All six participants in the coarser-scale condition and only one participant in the finer-scale condition showed correlations significantly different from zero ($n = 36$). The averaged correlation was .472 in the coarser-scale condition, and .076 in the finer-scale condition, with a difference of .396 ($t(5) = 7.906$, $p = .001$, two-tailed). This finding was consistent with the difference in accuracy between two conditions, supporting the speculation that participants were more likely to make a random guess in the finer-scale condition.

Table 2 Comparing Coarser-scale and Finer-scale

	Coarser-scale	Finer-scale	Difference
Accuracy	76.4% (8.61)	56.0% (8.05)	20.4%
Consistency	.472 (.060)	.076 (.094)	.396
RT	664.1ms (233.75)	798.4ms (298.93)	-123.2ms

N = 6. Standard deviations were listed in parentheses. All three comparisons were significant $p < .05$ (two-tailed).

The response time was also significantly different between two conditions. The averaged RT was 664.1ms in the coarser-scale condition and 798.4ms in the finer-scale condition, and the former was 123.2ms faster that the latter ($t(5) = -3.118$, $p < 0.05$, two-tailed). Since the response time was recorded as the latency after the onset of comparison numbers rather than the onset of poker cards, it is not clear whether probability estimation or number comparison, or both, produced the difference. Previous studies found that the time to make magnitude comparisons decreases linearly as the numerical distance between two numbers (e.g., Moyer & Landauer, 1967). Nevertheless, the large RT difference in our experiment indicated that the task of probability estimation and comparison as a whole might take more efforts in the finer-scale condition than in the coarser-scale condition.

ERP Results We rejected trials with voltages exceeding ± 100 μV. The remaining trials were segmented then averaged in synchrony with stimulus onset (display of poker cards) in a window of 1100ms (100ms before and 1000ms after stimulus onset), digitally transformed to an average reference, band-pass filtered (0.5 to 20 Hz), and corrected for baseline over 100ms before stimulus onset. Experimental conditions (Coarser vs. Finer) were compared on the 10 central-parietal electrodes by a repeated-measure ANOVA. We found significant ERP difference between two conditions at 400 ± 20ms (peak values) following stimulus onset (Greenhouse-Geisser $F(1,5) = 12.511$, $p < 0.05$), where the finer-scale condition yielded more positive voltages over parietal electrodes. Figure 3 shows the wave forms of electrode CP1 (GSN 38) and the voltage difference map (finer-scale minus coarser scale) by spherical spline interpolation. The comparison between the left and right hemispheres over the parietal area was not significant (Greenhouse-Geisser $F(1,5) = 5.339$, $p = .127$). At the current stage of the study, we have not found significant voltage differences over other brain areas. Furthermore, examinations on latency did not reveal any significant differences between two experimental conditions.

Examination of the waveforms showed that the ERP difference occurred after the P300 component, when participants were viewing identical displays and had not yet received the comparison stimuli. The P300 and its sub-components p3a and p3b have been considered as a process that indexes the ensuing memory storage operations, as

P300 amplitudes were found related to memory of previous stimulus presentations (e.g., Fabiani, Karis, & Donchin, 1990; Johnson, 1995; Paller, McCarthy, & Wood, 1988; for a recent review, see Polich, 2003). Our results are also consistent with those of Dehaene and colleagues (Dehaene et al., 1999; Naccache & Dehaene, 2001), who showed that 200-400ms after the stimulus onset is critical to distinguish among different numerical operations with distinctive semantic implications. Based on this observation, we speculate that the difference in ERPs in our experiment probably was due to different working memory load. Specifically, when judging the winning chance of a certain poker hand, other poker hands (either from previous trials or by temporary construction) were used as anchors to build categories on the internal representation. Fewer anchors were needed in the coarser-scale condition. On the contrary, the finer-scale condition demanded more effort because more hands needed to be considered at the same time. This speculation is consistent with the participants' oral report after the experiment. For example, one participant reported that he had to think of more hands with "nearby" rankings in the finer-scale condition.

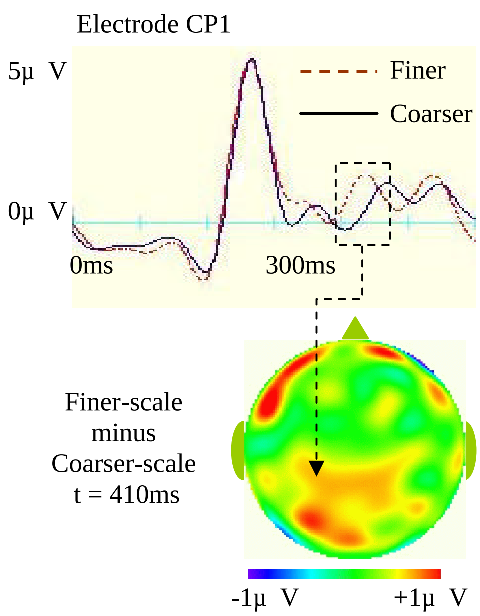

Figure 3

General Discussion

In sum, the findings in our experiment were consistent with our hypotheses of the internal representation of a coarser scale in intuitive judgment of event likelihood. The analyses of behavioral data suggested that if participants were only able to distinguish event likelihood with a coarser scale, a large portion of errors would be produced when they were forced to estimate probabilities with a finer scale. This kind of errors was manifested in both judgmental accuracy and consistency. The low accuracy level (close to random guess) and the low consistency level (close to zero) in the finer-scale condition indicated that participants were not able to distinguish winning probabilities of poker hands in the increment of 10%. Therefore, if the internal scale indeed had a limited number of categories, most likely this number was not greater than 10. It is interesting to point out that in our experiment, there were 36 different hands at 18 equally distributed levels of winning probabilities. If these hands were presented *externally* at the same time in the order of their rankings, it is reasonable to assume that any participant can report probability values accurate to the 5% increment by anchoring and adjustment. Nevertheless, their poor performance in the finer-scale condition indicated that the number of anchors that can be processed *internally* was quite limited. Furthermore, the ERP difference occurred when participants were viewing identical displays, indicating that participants followed the experimental instruction in forming their estimates at different levels of refinement. And estimating by a finer scale appeared to require more computational effort.

Note that the present study is still at its preliminary stage and there are many questions left to be answered. For example, more replications are needed and the ERP analyses can be extended such as comparisons over other brain areas and source localization. It would also be interesting to test someone who is an expert at poker and to see whether the performance would be better, especially in the finer-scale condition. Another example is that the model of a coarser internal scale can be further examined by manipulating the external representations, as previous studies indicated the important roles of the interaction between internal and external representations in human numerical cognition (e.g., Zhang & Norman, 1995; Zhang & Wang, in press). Upon further experiments and analyses, we believe that the present study will provide a better understanding of human intuitive probabilistic judgment.

Acknowledgements

This research was supported in part by the postdoctoral fellowship from the W. M. Keck Center for Computational and Structural Biology of the Gulf Coast Consortia for the first author.

References

Chapman, G. B., & Johnson, E. J., (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 120-149). New York: Cambridge.

Chase, V. M., Hertwig, R., Gigerenzer, G., (1998). Visions of rationality, *Trends in Cognitive Sciences, 2(6)*, 206-214.

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.

Cox, III, E. P., (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, *17*, 407-422.

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

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L., (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience, 21*, 355-361.

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

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 249-267). New York: Cambridge University Press.

Fabiani, M., Karis, D., & Donchin, E. (1990). Effects of mnemonic strategy manipulation in a von Restorff paradigm. *Electroencephalography and Clinical Neurophysiology, 75*, 22-35.

Johnson, R. (1995). Event related potential: Insights into the neurobiology of memory systems. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology*, Vol. 10, (pp. 135-163). Amsterdam: Elsevier.

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

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582-591.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*, 263-269.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126(3)*, 278-287.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology*, Vol.2, (pp. 83-115). Chichester, England: Wiley.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability*. New York: Wiley

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

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

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). Heuristics and biases: *The psychology of intuitive judgment*. New York: Cambridge.

Goldstein, W. M, & Hogarth, R. M. (1996). Judgment and decision research: Some historical context. In Goldstein, W. M & Hogarth, R. M. (Eds.), *Research on judgment and decision making: Currents, connections, and controversies* (pp. 3-65). New York: Cambridge University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519-1520.

Naccache, L., & Dehaene, S. (2001). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex, 11*, 966-974.

Osherson, D., Lane, D., Hartley, P., & Batsell, R. (2001). Coherent probability from incoherent judgment. *Journal of Experimental Psychology: Applied, 70(1)*, 3-12.

Paller, K. A., McCarthy, G., & Wood, C. C. (1988). ERPs predictive of subsequent recall and recognition performance. *Biological Psychology*, *26*, 269-276.

Pesenti, M., Thioux, M., Seron, X., Volder, A. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: A PET study. *Journal of Cognitive Neuroscience*, *12(3)*, 461-479.

Polich, J. (2003). Theoretical overview of P3a and P3b. In J. Polich (Ed.), *Detection of change: Event-related potential and fMRI findings* (pp. 83-98). New York: Kluwer Academic Publishers.

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

Stanescu-Cosson, R., Pinel, P., Moortele, P., Bihan, D. L., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain*, *123*, 2240-2255.

Sun, Y. & Tweney, R. D. (2002). Detecting the local maximum: The adaptive significance of a simple heuristic. In C. Schunn & W. Gray (Eds.), *Proceedings of the twenty-fourth annual conference of the Cognitive Science Society*, (pp. 856-860). Mahwah, NJ: Lawrence Erlbaum Associates.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, *87*, 154-163.

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

Wang, H., Johnson, T. R., & Zhang, J. (in press). The order effect in human abductive reasoning: An empirical and computational study. *Journal of Experimental and Theoretical Artificial Intelligence*.

Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition*, *57*, 271-295.

Zhang, J., & Wang, H. (in press). The effects of external representations on numerical tasks. *Quarterly Journal of Experimental Psychology*.