

Optimal Inference and Feedback for Representational Change

Yun Tang (tang.162@osu.edu)

Department of Psychology, The Ohio State University,
1835 Neil Ave, Columbus, OH 43210 USA

Christopher J. Young (young.1202@osu.edu)

Jay I. Myung (myung.1@osu.edu)

Mark A. Pitt (pitt.2@osu.edu)

John E. Opfer (opfer.7@osu.edu)

Abstract

Knowledge representations are central to many cognitive processes, and how these representations change is a central issue in learning and cognitive development. Here we developed and implemented a Bayesian inferential procedure to detect and elucidate representational change in numerical estimation. The proposed procedure of an adaptive numerical experiment both infers a learner's representation and predicts the feedback that is likely to induce representational change. We provide an application of this procedure using simulated subjects and demonstrate its effectiveness in inferring representational state and inducing change.

Keywords: representational shift; numerical estimation; adaptive experiment; Bayesian inference.

Introduction

Knowledge representations play a large role in cognitive processes such as learning, memory, and problem-solving (Markman, 1999), and a central problem in cognitive development concerns how representations change with age and experience (Carey, 1985; Dixon & Bangert, 2002; Siegler & Opfer, 2003). A striking example of representational change occurs in developing numerical magnitude representations. These representational changes are apparent across a wide range of tasks where numbers are quantified along a range, whether by categorizing numbers by magnitude (Opfer & Thompson, 2008), estimating numerosity (Booth & Siegler, 2006), measurements (Booth & Siegler, 2006), or positions of numbers on number lines (Dehaene, Izard, Spelke, Pica 2008; Siegler & Opfer, 2003).

Studies on development of numerical representations typically find that young children initially estimate numerical magnitudes to increase logarithmically with actual value before later learning the decimal system (Siegler & Opfer 2003, Booth & Siegler 2004; Opfer & Thompson 2007). This change is interesting theoretically because the logarithmic representation is implicit in speeded magnitude comparisons (Moyer & Landauer, 1967) and generation of random numbers (Banks & Hill, 1974) despite explicit judgments of numerical magnitude. This shift is also widespread across cultures, occurring relatively early in cultures that emphasize children's mathematical education

(Siegler & Mu, 2008) and delayed in cultures that lack formal schooling (Dehaene, Izard, Spelke, & Pica, 2008). Recent evidence also suggests that this representational shift can be induced in situ by providing examples (Izard & Dehaene 2007; Opfer & Siegler, 2007). That is, feedback on a few key numbers that are highly discrepant between logarithmic and linear functions causes rapid and broad adoption of linear representations (Opfer & Siegler, 2007).

Ideally feedback should take into account a child's current and target representational states. To do so, one must first infer, from a few noisy examples, the model that best describes the child's perception of numerical magnitude. This inference may be viewed as a model selection problem in which candidate models are evaluated and compared for their ability to capture the regularities underlying the data (Pitt & Myung, 2002). With the underlying representation having been inferred, one is now in a position to determine feedback that is most likely to induce representational changes in learners. This latter perspective proposes hypotheses for the ideal training regimen; feedback given to a child will be the most effective when it maximally discriminates between a logarithmic and linear representation while tracking the learner's current representation. These ideas can be formalized in a statistical framework, which is described in detail in a later section. This formal approach should have benefits to the theoretical questions that motivate research on the shift in numerical estimation, i.e. what is the path and source of change in numerical estimation abilities? We will be able to measure more precisely what about a child's representation changes to and what types of feedback are most likely to elicit it. The fruits of this approach could lead to the introduction of more effective teaching and training regimens.

In the present paper we propose a procedure that both (1) adaptively infers a learner's most likely representation and (2) predicts the feedback that will most likely induce representational shifts through what we call a cognitive tutor. We will demonstrate how this procedure is performed using computer simulations with information drawn from previous experimental data. We will also show the advantages of this procedure over traditional training studies in efficiency and the likelihood of inducing change.

Given our present focus on simulations of the above procedure, the purpose of this simulation study is three-fold. First, before implementation in experimental settings, it is necessary to run simulations to check the performance and accuracy of the method. Second, simulations could demonstrate the advantages of the cognitive tutor over the traditional paradigm. Finally, we are able to generate hypothesis for later experiments from simulation results. We use the topic of numerical estimation as a running example, and then discuss the potential to transfer the technology to other domains.

Adaptive Numerical Experiment

For representational shift problems, specifically in the domain of numerical representation, we propose an adaptive numerical experiment which infers the representation and performs the role of cognitive tutor. The procedure takes a perspective of model selection and distinguishes between the following models:

$$y_i = ax_i + b + e_i \quad (i = 1, \dots, n) \quad (1)$$

$$y_i = a \log x_i + b + e_i \quad (i = 1, \dots, n) \quad (2)$$

where x denotes the presented stimuli, y denotes the perceived numerical magnitude, and e is a normally distributed error with mean 0 and standard deviation σ .

In the experiment, we follow the paradigm used in Opfer and Siegler (2007), which shows the importance of choosing feedback. The same Number-Line Task is used as the numerical estimation task in our experiment. In each experiment trial, the child is shown a number between 0-100 or 0-1000 and is asked to estimate its position on a line.

The experiment is split into three sessions, as illustrated in Figure 1, and mirrors previous number line studies. In the pre-test session (Session 1), the Number-Line Task is performed to infer the child's existing representation model; each trial children are shown a number and asked to estimate its corresponding position on a line. Next in the feedback session (Session 2), children respond as in pre-test, but after each response are shown the (correct) linear position of each number. The post-test session (Session 3) is similar to the pre-test session, which examines whether any shift occurred in the child's representation model, with no feedback provided.

The proposed adaptive numerical experiment applies the Adaptive Design Optimization (ADO) method and reorganizes the three sessions into two processes, the adaptive inference and the adaptive tutoring. In what follows we define the two processes and describe how ADO works and how it is incorporated into the processes.

Adaptive Inference Process

The adaptive inference process (AIP) takes place in the pre-test session and infers a child's most likely representation model (e.g. linear). It conducts a series of experiment trials and presents the numerical stimuli sequentially. Within each trial, the observed response is analyzed and the next stimulus is provided based on the

analysis. It is adaptive in that it tailors the test procedure to individual state from trial to trial. Consequently, it obtains sufficient evidence to make inference within the fewest possible trials.

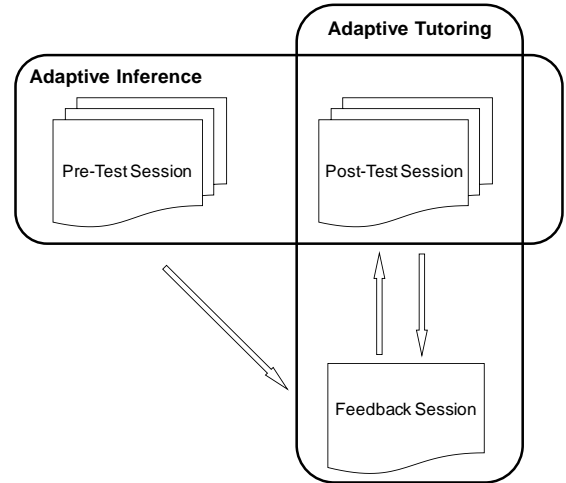


Figure 1: General structure of adaptive numerical experiment consisting of the adaptive inference and the adaptive tutoring processes.

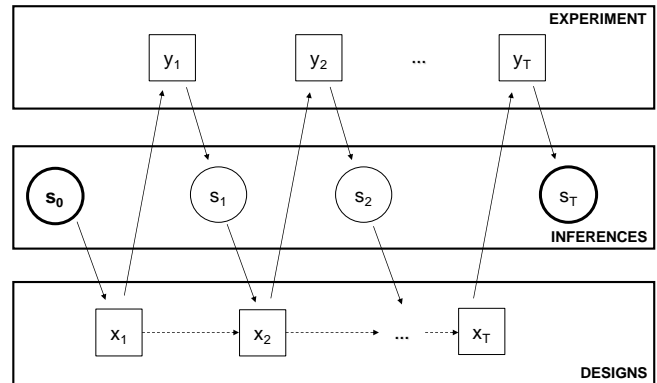


Figure 2: Flowchart of ADO process including repeated sessions of design optimization (designs), data collection (experiment), and model updating (inferences).

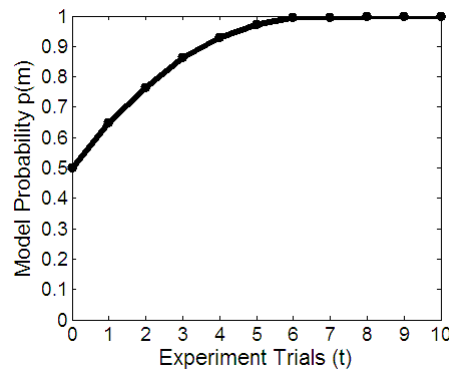


Figure 3: A typical curve of model probability change in ADO experiments.

The adaptive choice of numerical stimuli is formally done via experiment design optimization methods, where the numerical stimuli are the designs of interest. The idea of design optimization in this task is to find a numerical stimulus that is the most informative in distinguishing among alternative representational formats (i.e., logarithmic vs. linear). This method of adaptive design optimization (ADO) is developed and performed in a Bayesian framework (Myung & Pitt, 2009). In ADO, design optimization (designs), data collection (experiment), and model updating (inferences) are repeatedly performed, as illustrated in the flowchart in Figure 2. In the process, x denotes the numeric value presented to the child and is the design variable to be optimized. The symbol y denotes the child's response, and s denotes the current inference about the child's underlying representation state, such as the relative likelihood of candidate models and their parameters, which are formally defined later. The numbers the child sees in the session are updated trial by trial along the experiment.

The ADO process is performed as follows. At the beginning, the experimenter has some prior information s_0 about the child's model, from which the initial number x_1 is drawn and the response y_1 is observed. With s_0 and y_1 , the posterior s_1 is obtained by Bayes theorem. For the next trial, s_1 serves as the prior and the above process is repeated. The process continues until the model information s_T after T trials meets certain stopping criterion. Such an adaptive approach bases the later designs upon previous experimental results and makes better use of individual data. Hence, it is more efficient compared to the traditional manner of using the same designs for every individual. Figure 3 shows a typical curve of model probability obtained from an ADO simulation. It indicates that the predicted model probability of the true underlying model reaches as high as .9 within four trials. To summarize, ADO-embedded adaptive inference process could find the optimal designs (i.e. numerical values to estimate) that tailor to individual state, thus could permit efficient inference from the results.

Adaptive Tutoring Process

After inferring the child's representation model through AIP, we may know that the child uses some undesired logarithmic or linear model. The next concern is to find appropriate feedback stimuli that will be most likely to induce representational shift. For this purpose, we combine the feedback session and the post-test session to form what we call the adaptive tutoring process (ATP). Design optimization methods are also applied in ATP. In the feedback session, the choice of the feedback stimuli is optimized in order to teach the child most effectively. For this purpose, we make the assumption that the effectiveness of the design is determined by the maximum discrepancy between the child's model and the target model (e.g., an accurate line $y_i = x_i$). After the optimal feedback stimulus is found and provided to the child, ATP moves to the post-test session. The post-test session infers the child's model again and checks if he has changed the model. If the child

retains a logarithmic model or changes to an undesired linear model (e.g. a linear model with slope smaller than .5), the feedback and post-test sessions are repeated until the child has acquired the target model. Generally speaking, adaptive inference is also performed within the adaptive tutoring process.

The adaptive tutoring process starts from the information s_T obtained at the end of the adaptive inference process. In determining the numbers to be used for teaching, our assumption is that the most informative feedback stimuli for the child lie in the region where the target model and the child's current representation model have the largest discrepancy. The target model is assumed as a fixed, correct model. Hence, we are not adapting to the child's representation states, but are optimizing to the difference between the child's current status and the target model. Formally, we are maximizing the informativeness of the feedback stimulus described as the discrepancy between its true value and its value in the child's representation. The child is tested with the optimal feedback and is corrected with the true position. Then the experimenter obtains the updated information about the child's numerical representation model using the same process as in AIP. The updated information can be used to find the next optimal feedback stimulus, if necessary. The process runs back and forth until the child has shown acquisition of the target model by giving accurate linear responses to the numerical stimuli. In all, the adaptive tutoring process tailors to the child's learning progress and provides a way to combine optimal teaching and progress verification.

Bayesian Framework of Design Optimization

In this section, we provide a brief description of the ADO framework implemented in this paper. For fuller technical details and applications, the reader is directed to Myung and Pitt (2009) and Cavagnaro, Myung, Pitt and Kujala (2010). In ADO, each experimental design is assigned a utility describing the value of a hypothetical experiment with that design. It is analogous to choosing among a set of gambles whose payoff is determined by the risks and rewards of each type of gamble. The set of all possible designs that could be used in a given experiment consist of the design space (Amzal, Bois, Parent, & Robert, 2006; Pitt & Myung, submitted). The goal of ADO is to search the entire design space and find the most informative design(s).

The problem of design optimization is formally expressed as finding an optimal design d^* over the design space, which maximizes the expected utility function $U(d)$. $U(d)$ typically takes into consideration of all unknown but possible conditions. If multiple models are plausible for describing the underlying process in an experiment, $U(d)$ could be defined as:

$$U(d) = \sum_{i=1}^K p(m_i) \iint u(d, m_i, \theta_{m_i}, y) p(y | \theta_{m_i}, d) p(\theta_{m_i}) dy d\theta_{m_i} \quad (3)$$

In the above equation, m_i ($i = \{1, \dots, K\}$) is one of K models under consideration, d is a design, y is the outcome of an experiment with design d under model m , θ_m is the

parameter of model m , and finally, $u(d, \theta_m, y)$ is the “local” utility function of design d , parameter θ_m and experimental outcome y . In general, $U(d)$ represents the expected value of local utility functions in which the expectation is taken over all possible models and their parameters and over all possible experimental observations given the models and parameters.

In *adaptive* design optimization, the optimization of $U(d)$ is repeated over a series of experimental stages. At each stage, the model and parameter priors, $p(m)$ and $p(\theta_m)$, are updated upon the specific outcome observed in an actual experiment carried out with the optimal design d^* . This updating is performed via Bayes rule and Bayes factor calculation (Gelman, Carlin, Stern & Rubin, 2004).

Simulations

Pre-test Simulations and Results

In this section, we describe the computer simulations that demonstrate the performance and advantages of the adaptive numeric estimation experiment. The purpose of conducting simulations is to guarantee that the processes work as expected, as well as to show the efficiency of the methodology.

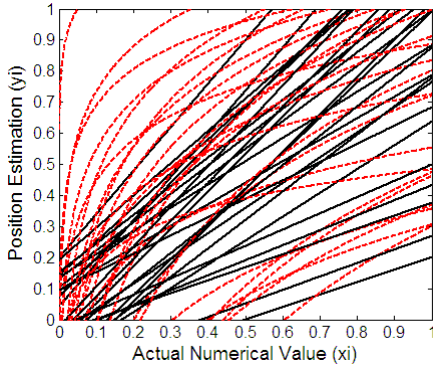


Figure 4: Sample curves of linear (black solid lines) and logarithmic (red dashed curves) models randomly generated from the priors.

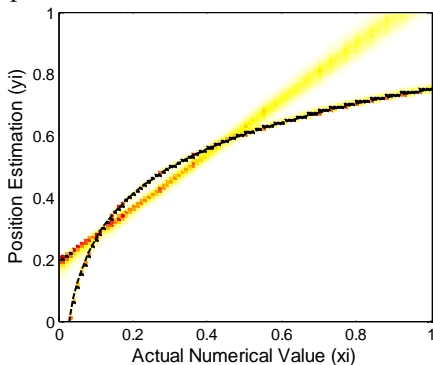


Figure 5: Prediction density scatter plot of linear and logarithmic model predictions at the end of the pre-test. The darkness of each dot indicates the probabilities of a response y given the presented number x . Black dots indicate the highest probabilities and the yellow dots indicate the lowest probabilities.

In order to run the simulations, we first chose the priors on the basis of previous experiment data and experts’ beliefs, so that the priors covered a reasonable range of numerical representation models. Several data sets (e.g. Opfer & Siegler, 2007, Siegler & Opfer, 2003) were fitted and the parameter ranges of the models were obtained. Uniform priors over the parameter ranges were then used for intercept, slope, and error variance. Figure 4 shows a sample of possible models under the priors, in which the linear models and logarithmic models are mixed with each other. It also suggests the difficulty of depicting intuitive designs for distinguishing between the two sets of models.

The simulation first implemented the pre-test session with the above priors. The data-generating model, which was assumed to be the child’s true model in the simulations, took the following logarithmic form:

$$y_i = 0.21 \cdot \log x_i + 0.75 + e_i, \quad e_i \sim N(0, 0.005^2)$$

Within each simulation, we ran 10 trials (number of trials fixed for convenience purposes) of the Number-Line Task in the pre-test session. Results showed that after 6 trials, we had already obtained sufficient evidence to conclude that the logarithmic model was over 90% likely to be the data-generating model. Meanwhile, we also narrowed down the range of model parameters as shown in the prediction density scatter plot in Figure 5. The darkness of each dot indicates the probabilities of a response y given the presented number x . Figure 5 shows that the predictions from possible linear models are more widely spread than the predictions from possible logarithmic models. It suggests that the predictions from the logarithmic model posteriors are highly concentrated and have higher probabilities, which provides strong evidence that the true model takes a logarithmic form.

Feedback and Post-test Simulations

After the pre-test session, we simulated the adaptive tutoring process. The first step was to choose an optimal feedback stimulus that maximized the discrepancy between the target model and what we knew about the child’s existing model. Formally, the utility of the feedback design accounted for the prediction probabilities of both models, as well as the parameter range of both models. For the specific simulated learner, the optimal feedback design was found at $x = 0.354$. That is, the child would be most “surprised” for this stimulus when he sees the difference between his response and the correct answer. Figure 6 shows the location of the optimal feedback and its relationship with the child’s model and the target model.

To simulate the post-test session, we needed to assume a learning mechanism that caused the representational shift and generated the post-test experiment results. An intuitive assumption was a conservative learning mechanism in which a child learner made the smallest change to accommodate the feedback. Suppose the child could change to any models within the range of the priors. Among these models, there were a subset of linear models and a subset of logarithmic models that were consistent with the learned

feedback. A conservative learner would estimate the amount of overall discrepancy between these candidate models and the current model and choose the one that has the smallest discrepancy. That is, the conservative learning mechanism assumed the child to be an ideal learner. To demonstrate another plausible mechanism, we also assumed a less ideal learner, the model-conservative learner. The model-conservative learning mechanism assumed that the child only considered a subset of logarithmic models that were consistent with the learned feedback and chose one that required the smallest change from the previous model. In both mechanisms, the winning model was used as the data-generating model for the post-test session. Figure 7 shows representational shifts of the two hypothesized learners. After learning the optimal feedback, the conservative learner changes to a linear model $y_i = 0.758 \cdot x_i + 0.086 + e_i$, and the model-conservative learner changes to another logarithmic model $y_i = 0.218 \cdot \log x_i + 0.580 + e_i$. The two models intersect at the point of optimal feedback because they both accommodate the feedback.

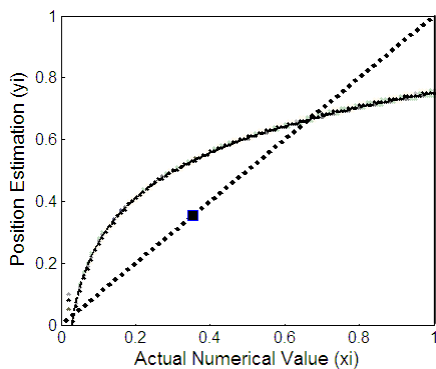


Figure 6: Optimal feedback for the simulated learner indicated by the square at $x = 0.354$. The prediction density scatter plot shows the inference of the child's representation at the end of the pre-test session. The dotted line shows the target model ($y_i = x_i$).

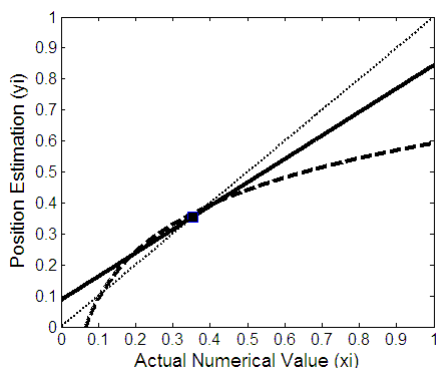


Figure 7: Predicted representational shift to the linear model (solid line) and the logarithmic model (dashed curve) caused by the two learning mechanisms. Both models intersect with the target model at the feedback (the square).

The post-test session simulation started from the same priors used for the pre-test session (shown in Figure 4). It was because the data-generating model had changed and the posterior information from the pre-test session was no longer valid. For convenience purpose, we simulated 5 trials of Number-Line Task in the post-test session. For the conservative learner, there was sufficient evidence to conclude that linear model was over 90% likely to be the data-generating model after 4 trials. For the model-conservative learner, it took 5 trials. The range of parameter estimates for the data-generating model was also narrowed down at the end of the post-test session. Hence, results from the post-test simulations showed that the post-test session made quick and reliable inferences about the new data-generating model.

In general, simulation results of the pre-test, feedback, and post-test sessions demonstrated the validity and the efficiency of the adaptive numerical experiment. We further discuss its practical applications and theoretical implications in the next section.

Discussion

Previous feedback studies have demonstrated that providing children with data that is incommensurate with their current numerical representation can promote a representational shift. In the current paper we improved upon this design using an adaptive design optimization procedure to perform an adaptive-inference, adaptive-tutoring process. This process infers the most likely dominant numerical representation and provides the optimal feedback to elicit a shift to an accurate linear representation. We simulated this process for a logarithmic learner using parameters from previous empirical experiments. Finally we predicted the learner's updated numerical representation based on two possible learning mechanisms.

We established the plausibility of the algorithm for the problem at hand. The adaptive design optimization procedure was able to infer the data generating function in each simulation by optimizing across the design space. The procedure was more efficient than traditional feedback studies in inferring the simulated child's representational state in a few trials. This efficiency in turn suggests that a shorter pre-test phase is less likely to reinforce the learner's initial representation. Shorter testing and feedback phases also provide obvious benefits to both experimentation and real world application for testing children; fewer trials reduce the overall attentional costs to children and thereby reduce the influence of attention-related noise in their responses.

The adaptive tutoring process also proved useful in determining optimal feedback. Feedback points have previously been chosen to maximize the discrepancy between an ideal logarithmic and linear function (Opfer & Siegler, 2007), while our cognitive tutor chooses personalized feedback based on the individual learner's most likely logarithmic or linear representation. This generates very informative results about the ideal feedback points.

The magnitudes chosen by the adaptive tutor are approximately 30% of the range for a simulated learner based on the parameters of children from previous studies. They are near to the previously chosen points (15% of the range), but are clearly not the same. These optimal feedback points may prove to vary widely in actual children, highlighting the need for the adaptive tutoring process to control for individual differences in representations.

The adaptive numeric estimation experiment clearly needs to be run on children to determine its external validity, which we plan to carry out. Nevertheless, we were able to use the adaptive experiment to accurately infer the representational state of a simulated learner. A byproduct of this process was the implementation of two potential learning mechanisms to test the end-state representation of the simulated learner. The conservative and model-conservative learning mechanisms were used to produce quantitative predictions. A conservative model that uses optimal feedback to adjust parameters and the model form with the least amount of change showed a shift to a more accurate linear function with parameters near to the ideal model. The model-conservative mechanism resulted in a preserved logarithmic function with an overall decrease in the model parameters.

If these results can be extended to children, they would support a perspective that learner will behave as a modeler and update his dominant representation with ideal feedback. We might then further test whether the child learner is engaging in Bayesian learning; specifically whether the different learning mechanisms can be seen as a variation in the learner's likelihood ratio. Conservative learning asserts equal likelihood to the representations, while model-conservative learning gives weight only to the dominant representation. These may be plausible mechanisms of cognitive change based on culture and the strength of each representation, with emphasis on mathematical education directly affecting the learner's likelihood ratio of a linear representation.

Adaptive inference of the probability that a learner is linear or logarithmic in representation and an adaptive tutor function that maximizes the effect of feedback are necessary to understand the learner's representation which might apply to many types of representations in diverse areas. The process could easily be extended to similar numerical estimation tasks that use a variety of presented numerical stimuli to determine perceived magnitude. It is possible to extend this design to other areas in which representational shifts are seen, whether to determine children's past tense verb use and predict errors in overgeneralization (Marcus, 1995) or function learning to predict attention to relevant cues (Kruschke 1996). The adaptive design optimization procedure is of obvious use as a means of better modeling the learner and refining training.

Acknowledgment

This research is supported in part by the NIH Grant R01-MH57472 to JIM and MAP.

References

- Amzal, B., Bois, F. Y., Parent, E., & Robert, C. P. (2006). Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical Association, 101*(474), 773-785.
- Banks, W. & Hill, D. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology, 102*(2), 353-376.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation, 22*(4), 887-905.
- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian Indigene Cultures. *Science, 320*, 1217-1220.
- Dixon, J., & Bangert, A. (2002). The prehistory of discovery: Precursors of representational change in solving gear system problems. *Developmental Psychology, 38*(6), 918-933.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis. (2nd ed.)*. Boca Raton, FL: Chapman and Hall/CRC.
- Kruschke, J. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8*(2), 225-247.
- Marcus, G. (1995). The acquisition of the English past tense in children and multilayered connectionist networks. *Cognition, 56*, 271-279.
- Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Moyer, R. & Landauer, T. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519-1521.
- Myung, J. I. & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review, 116*(3), 499-518.
- Opfer, J. & Siegler, R. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169-195.
- Opfer, J. & Thompson, C. (2008). The trouble with transfer: Insights from microgenetic changes in the representation of numerical magnitude. *Child Development, 79*(3), 788-804.
- Pitt, M. A. & Myung, I. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6*, 421-425.
- Pitt, M. A. & Myung, J. I. (submitted). Designing a better experiment. *Manuscript submitted for publication*.
- Siegler, R. & Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychological Science, 19*(8), 759-763.
- Siegler, R. & Opfer, J. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science, 14* (3), 237-43.