Title
A Hierarchical Adaptive Approach to the Optimal Design of Experiments

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

Authors
Kim, Woojae
Pitt, Mark
Lu, Zhong-Lin
et al.

Publication Date
2014

Peer reviewed

A Hierarchical Adaptive Approach to the Optimal Design of Experiments

Woojae Kim[1] (kim.1124@osu.edu), Mark Pitt[1] (pitt.2@osu.edu), Zhong-Lin Lu[1] (lu.535@osu.edu), Mark Steyvers[2] (mark.steyvers@uci.edu), Hairong Gu[1] (gu.124@osu.edu), & Jay Myung[1] (myung.1@osu.edu)

[1]Department of Psychology, Ohio State University, Columbus, OH 43210 USA
[2]Department of Cognitive Sciences, University of California, Irvine, CA 92697 USA

Abstract

Experimentation is at the core of research in cognitive science, yet observations can be expensive and time-consuming to acquire. A major interest of researchers is designing experiments that lead to maximal accumulation of information about the phenomenon under study with the fewest possible number of observations. In addressing this challenge, statisticians have developed adaptive design optimization methods. This paper introduces a hierarchical Bayes extension of adaptive design optimization that provides a judicious way to exploit two complementary schemes of inference (with past and future data) to achieve even greater accuracy and efficiency in information gain. We demonstrate the method in a simulation experiment in the field of visual perception.

Keywords: optimal experimental design, hierarchical Bayes, mutual information, visual spatial processing.

Introduction

Experimentation advances cognitive science by providing quantified evidence for evaluating and testing theories. From the standpoint of information theory (Cover & Thomas, 1991), one should design an experiment that across trials seeks to gain as much information as possible from the cognitive process under study. For example, experiments often require choosing levels of an independent variable (e.g., stimuli of different sizes or intensities). These choices impact the informativeness (or quality) of the resulting data, which in turn impacts what can be concluded about the issue of interest.

As a concrete example, consider an experiment in visual psychophysics in which one is interested in estimating a viewer's ability to see fine detail. When the sensitivity is measured with stimuli that vary not only in their contrast but also in their spatial frequency, the measurements form a contrast sensitivity function (CSF; Figure 1). A CSF characterizes a person's vision more accurately than traditional visual acuity measurements (i.e., using an eye chart) and is often useful for detecting visual pathologies (Comerford, 1983). However, because the standard methodology (e.g., staircase procedure) can require many hundreds of trials for accurate estimation of the CSF curve, it is a prime candidate for improving information gain, and Lesmes, Lu, Baek and Albright (2010) developed such a method. In each trial of the experiment, the to-be-presented stimulus is chosen such that it maximizes information gain by adaptively taking into account what has been learned about the participant's performance from past trials.

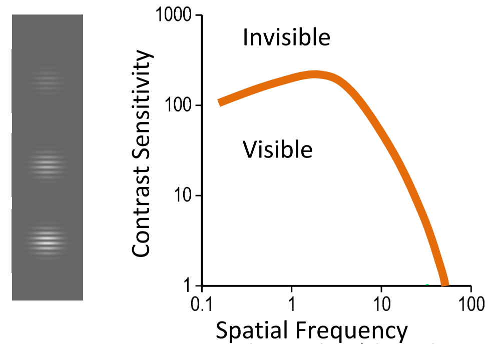

Figure 1: Examples of stimuli (left) and a typical contrast sensitivity function (right). Sinewave gratings with varying contrast and spatial frequency are used to measure a person's CSF.

Their procedure was one implementation of algorithm technology being developed in the burgeoning field of design optimization (Atkitson & Donev, 1992; Amzal, Bois, Parent, & Robert, 2006). Called *adaptive design optimization* (ADO; e.g., DiMattina & Zhang, 2008; Cavagnaro, Tang, Myung & Pitt, 2009), this method capitalizes on the sequential nature of experimentation, making each new measurement using the information learned from previous measurements of a subject so as to achieve maximal information about the cognitive process under study.

As currently used, ADO is tuned to optimizing a measurement process at the individual participant level, without taking advantage of information available from data collected from other individuals or testing sessions. To illustrate the situation using the example of CSF estimation, consider a space of CSFs in Figure 2, a point in which represents an individual's measured CSF (e.g., the ones shown with arrows). Suppose that one has already collected CSFs from a group of participants (dots in the ellipse) and data are about to be collected from one more (triangle). Knowledge of what this person's CSF function might look like can be informed by the group data, and thereby expedite and improve data collection. Without the benefit of the group-level information, data collection would be more time-consuming because what constitutes a probable CSF would be unknown.

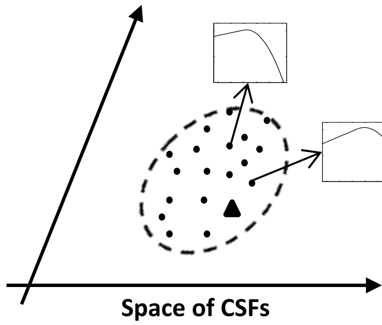

Figure 2: Illustration of a situation in which contrast sensitivity functions of a group of participants are measured in an experiment.

The purpose of the present investigation is to develop a general design optimization framework that extends the existing ADO methodology to incorporate the prior knowledge of the population characteristics that are available before the experiment to achieve even greater information gain. The proposed method, dubbed *hierarchical adaptive design optimization* (HADO; Kim, Pitt, Lu, Steyvers, & Myung, under review), is an integration of two existing techniques, hierarchical Bayesian modeling (HBM) and adaptive design optimization. We begin by reviewing these two components briefly, followed by a formal description of HADO and an application example.

Adaptive Design Optimization (ADO)

The literature on optimal experimental design goes back to the pioneering work in the 1950s and 1960s in statistics (Lindley, 1956; Box & Hill, 1967). The recent surge of interest in this field can be attributed largely to the advent of statistical computing, which has made it possible to solve more complex and a wider range of optimization problems. ADO is gaining traction in areas where data are costly to collect such as in neuroscience (e.g., DiMattina & Zhang, 2008) and drug development (e.g., Miller, Dette, & Guilbaud, 2007). In psychology and cognitive science, ADO is gradually recognized and applied in research such as retention memory (Myung, Cavagnaro, & Pitt, 2013), decision making (Cavagnaro, Pitt, Gonzalez, & Myung, 2013), psychophysics (Lesmes, Jeon, Lu, & Dosher, 2006), and development of numerical representation (Tang, Young, Myung, Pitt, & Opfer, 2010).

ADO is formulated as a Bayesian sequential optimization algorithm that is executed over the course of an experiment. The framework of ADO is depicted in the shaded area in Figure 3. On each trial of the experiment, on the basis of the present state of knowledge (prior) about the phenomenon under study, which is represented by a statistical model of data, the optimal design with the highest expected information gain is identified. The experiment is then carried out with the optimal design, and measured outcomes are observed and recorded. The observations are subsequently used to update the prior to the posterior using Bayes' theorem. The posterior in turn is used to identify the optimal design for the next trial of the experiment. These alternating steps of design optimization, measurement, and updating of the individual-level data model are repeated in the experiment until a suitable stopping criterion is met.

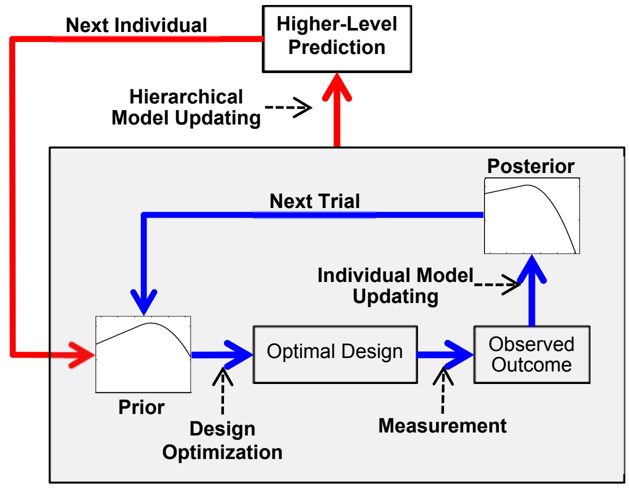

Figure 3: The framework of HADO. The shaded area represents the conventional ADO framework and the peripheral part is the hierarchical extension of ADO.

Hierarchical Extension of ADO

ADO optimizes designs using information available only on the individual participant level, without taking advantage of data collected from previous testing sessions. Hierarchical Bayesian modeling (HBM; Good, 1965; Lee, 2006) not only provides a flexible framework for incorporating this kind of prior information but is also well suited for being integrated within the existing Bayesian ADO. The basic idea of HBM is to improve the precision of inference (e.g., parameter estimation or power of a test) by taking advantage of statistical dependencies present in the data-generating structure that the individuals can be seen as being sampled from. From the Bayesian perspective, HBM can also be viewed as a justified way to form an informative prior from real, observed data. Thus, on a conceptual level, HADO may be best described as a method for integrating the ADO technique with previously available information about the population structure, in a Bayesian prior distribution, to maximize the efficiency of data collection even further.

This is illustrated in Figure 3, HADO being described by adding a loop over the conventional ADO. Being standalone, ADO starts with an uninformative prior for each individual participant. HADO extends ADO by modeling a higher-level structure across all individuals, which can be used as an informative prior for the next, new measurement session. In the other way around, the parameter estimate of a new individual helps update the higher-level structure.

Formulation of HADO

The implementation of conventional ADO, also a component of HADO, requires a statistical model defined as a parametric family of probability distributions, $p(y|\theta, d)$'s, which specifies the probability of observing an experimental outcome y given a parameter value θ and a design d. A prior for θ is assumed at the beginning of the experiment, and after each observation, the prior is updated by Bayes' theorem, and then it serves as the prior for the next experimental trial. For each trial, a design with the largest value of a pre-defined utility function is selected. The information-theoretic choice of a utility function is mutual information (Cover & Thomas, 1991), which in the current context is given by

$$U(d_t) = \iint \log \frac{p(\theta|y^{(1:t)}, d_t)}{p(\theta|y^{(1:t-1)})} p(y^{(t)}|\theta, d_t) p(\theta|y^{(1:t-1)}) \mathrm{d}y^{(t)} \mathrm{d}\theta, \quad (1)$$

where $y^{(1:t)}$ is the collection of past measurements made from the first to $(t-1)$-th trials, denoted by $y^{(1:t-1)}$, plus an outcome, $y^{(t)}$, to be observed in the current, t-th trial conducted with a candidate design, d_t.

To integrate HBM into ADO, a higher-level model of the parameters $p(\theta_{1:n}|\eta)$ is assumed (e.g. a multivariate normal density with parameters η) where $\theta_{1:n} = (\theta_1, \dots, \theta_n)$ is the collection of model parameters for all n individuals. The joint posterior distribution of the hierarchical model given all observed data is expressed as

$$\begin{aligned} p(\theta_{1:n}, \eta|y_{1:n}) &\propto p(y_{1:n}|\theta_{1:n}) p(\theta_{1:n}|\eta) p(\eta) \quad (2) \\ &\propto \left[\prod_{i=1}^{n} p(y_i|\theta_i)\right] p(\theta_{1:n}|\eta) p(\eta), \end{aligned}$$

where $p(\eta)$ is the prior distribution for the higher-level model's parameters η. If all parameters $\theta_{1:n}$ and η can be represented on a multidimensional grid and satisfy a certain condition (θ_i's are conditionally independent given η), values of $p(\theta_{1:n}, \eta|y_{1:n})$ in Eq. (2) can be easily calculated by dividing the values on each grid point by the summation of all such values (i.e., normalization). Then we can obtain the marginal distribution of η by integrating Eq. (2) over $\theta_{1:n}$ as

$$p(\eta|y_{1:n}) = \iint p(\theta_{1:n}, \eta|y_{1:n}) \mathrm{d}\theta_{1:n}. \quad (3)$$

The estimates $\hat{\eta}$ can be obtained as the expectation $\hat{\eta} = \sum_{g=1}^{G} \eta_g p(\eta_g|y_{1:n})$ where G is the number of the grid points of η. As such, an informative prior for a new participant is constructed by plugging $\hat{\eta}$ in the higher-level model $p(\theta|\hat{\eta})$, which substitutes the uninformative prior in the conventional ADO.[1]

[1] Approximation of the prior in more general cases (e.g., a point estimate $\hat{\eta}$ is considered too restrictive to represent the prior, or the grid size is too large to manage) is discussed in Kim et al. (under review).

For HADO to be adaptive, Bayesian updating for posterior distribution is performed recursively on two different levels. On the individual level, only the lower-level parameters θ are updated after each observation during an experiment. On the upper level, the higher-level parameters η are updated at the end of each experiment through HBM. The estimate of η is used to calculate the prior for the next measurement session.

Simulation Experiments

The benefits of HADO were demonstrated in simulated experiments in the domain of visual perception (an example used in the Introduction). Using the conventional ADO framework described earlier, Lesmes et al. (2010) introduced an adaptive version of the contrast sensitivity test called qCSF. Contrast sensitivity, $S(f)$, against spatial frequency f, was modeled using the truncated log-parabola with four parameters:

$$S(f) = \begin{cases} \gamma^{max} - \delta & \text{if } f < f^{max} - \frac{\beta}{2}\sqrt{\frac{\delta}{\log_{10} 2}} \\ \gamma^{max} - (\log_{10} 2)\left(\frac{f - f^{max}}{\beta/2}\right)^2 & \text{otherwise,} \end{cases} \quad (4)$$

where the four parameters are γ^{max}, the peak sensitivity, f^{max}, the peak frequency, β, the bandwidth, and δ, the low-frequency truncation level.

To demonstrate the benefits of HADO, the simulation study considered four conditions in which simulated subjects were tested for their CSFs by means of four different measurement methods.

Simulation Design

The two most interesting conditions were the ones in which ADO and HADO were used for stimulus selection. In the first, *ADO* condition, the qCSF method of Lesmes et al. (2010) was applied and served as the existing, state-of-the-art technique against which, in the second, *HADO* condition, its hierarchical counterpart developed in the present study was compared. If the prior information captured in the higher-level structure of the hierarchical model can improve the accuracy and efficiency of model estimation, then performance in the HADO condition should be better than that in the ADO (qCSF) condition. Also included for completeness were two other conditions to better understand information gain achieved by each of the two components of HADO: hierarchical Bayes modeling (HBM) and ADO. To demonstrate the contribution of HBM alone to information gain, in the third, *HBM* condition, prior information was conveyed through HBM but no optimal stimulus selection was performed during measurement (i.e., no ADO). In the fourth, *non-adaptive* condition, neither prior data nor stimulus selection was utilized, so as to provide a baseline performance level against which improvements of the other methods could be assessed.

The hierarchical model in the HADO condition comprised two layers. On the individual level, each subject's CSF was modeled by the four-parameter, truncated log-parabola described above. On the upper level, the generation of a subject's CSF parameters was described by a four-variate Gaussian distribution, along with the usual, normal-inverse-Wishart prior. While a more refined structure might be plausible (e.g., the population is a mixture of heterogeneous groups, or CSFs covary with other observed variables), the current hypothesis (i.e., individuals are similar to each other in the sense that their CSFs are normally distributed) was simple and sufficient to show the benefits of HADO.[2]

The ADO (qCSF) condition shared the same individual data model as specified in the HADO condition, but the variability among individuals was not accounted for by a higher-level model. Instead, each individual's parameters were given a diffuse, Gaussian prior. The HBM condition took the whole hierarchical model from HADO, but the measurement for each individual was made with stimuli randomly drawn from a prespecified set. Finally, the non-adaptive method was based on the non-hierarchical model in ADO (qCSF) and used random stimuli for measurement.

To increase the realism of the simulation, we used real data collected from 147 adults who underwent CSF measurements. The number of measurements obtained from each subject was more than adequate to provide highly accurate estimates of their CSFs. These estimates were taken and assumed to be underlying CSFs in this simulation study.

To compare the four methods, we used a leave-one-out paradigm, treating 146 subjects as being previously tested and the remaining subject as a new individual to be measured subsequently. We further assumed that, in each simulated measurement session, artificial data are generated from an underlying CSF (taken from the left-out subject) with one of the four methods providing stimuli. This situation represents a particular state in the recursion of measurement sessions shown in Figure 3; that is, the session counter is changing from n = 146 to n = 147 to test a new, 147th subject. Theoretically, the two-stage updating shown in Figure 3 may be used from the start of a large-scale experiment (i.e., from n = 1). However, because a large sample is needed for the higher-level structure to be accurately estimated, HADO can be applied with no significant loss of its benefit in a situation in which there are some previously collected data.[3]

To prevent idiosyncrasies of the simulation's probabilistic nature (due to simulated subjects' random responses) from misleading the interpretation of results, ten replications of the 147 leave-one-out experiments were run independently and results were averaged over all individual sessions (i.e. 10×147=1470 experiments were conducted in total). All required computations for individual-level design optimization and Bayesian updating (i.e., shaded area in Figure 3) were performed on a grid in a fully deterministic fashion (i.e., no Monte Carlo integration). The posterior inference of the higher-level model (i.e., outside the shaded region in Figure 3), also involved no sampling-based computation. This was possible because the higher-level model (i.e., Gaussian distribution) allowed for conditional independence between individuals so that the marginal posterior distribution in Eq. (3) could be evaluated as repeated integrals over individual θ_i's.

Results

Performance of the four methods of measurement was first evaluated with respect to information gain. The degree of uncertainty about the current, n-th subject's parameters upon observing trial t's outcome was measured by the differential entropy (extension of the Shannon entropy to the continuous case; see Cover & Thomas, 1991). Use of the differential entropy, which is not bounded in either direction on the real line, is often justified by choosing a baseline state and defining the observed information gain as the difference between two states' entropies. In the present context, information gain is defined as

$$IG_t(\Theta_0,\Theta_n) = H_0(\Theta_0) - H_t(\Theta_n) \qquad (5)$$

where $H_0(\Theta_0)$ is the entropy of the baseline belief about θ in a prior distribution so that $IG_t(\Theta_0,\Theta_n)$ may be interpreted as the information gain achieved upon trial t during the test of subject n relative to the baseline state of knowledge. For all the four methods, we took the entropy of the non-informative prior as $H_0(\Theta_0)$.

Shown in Figure 4 is the cumulative information gain observed with the four different methods. Each of the four curves corresponds to information gain (y-axis) in each simulation condition over 200 trials (x-axis) relative to the non-informative, baseline state (0 on the y-axis). The information gain measures were averaged over all 1,470 individual measurement sessions in each condition.

The results demonstrate that the hierarchical adaptive methodology (HADO) achieves higher information gain than the conventional adaptive method (ADO). The contribution of hierarchical modeling is manifested at the start of each session as a considerable amount of information (0.4) in the HADO condition (solid curve) than

[2] Benefit of using a non-normal, mixture distribution for modeling higher-level structure was also investigated with this application example. A non-parametric, kernel density estimation (KDE; Hastie, Tibshirani, & Friedman, 2009) technique was employed to capture a highly non-normal, multimodal distribution on the space of CSF parameters in a simulation setup. Results suggested that, with a low-dimensional model such as a CSF model, the advantage of using such a mixture distribution is not significant, producing only slightly higher information gain than HADO with a normal distribution used as higher-level structure.

[3] Although not presented due to a space limit, an additional simulation was conducted to see how this application of HADO performs when there is a small accumulation of data (e.g., n = 4, 10, 40). The results suggest that the Bayesian estimation of this particular hierarchical model is robust enough to take advantage of even a small sample of previously collected data. However, the effect of small n may depend on the model employed, suggesting that this observation would not generalize to all potential HADO applications (Kim et al., under review).

no information (zero) in the ADO condition (dashed curve). As expected, this is because HADO benefits from the mutual informativeness between individual subjects, which is captured by the higher-level structure of the hierarchical model and makes it possible for the session to begin with significantly greater information. As the session continues, HADO needs 43 trials on average to reach the baseline performance of the non-adaptive method (dotted, horizontal line) whereas ADO (qCSF) requires 62 trials. The clear advantage diminishes as information accumulates further over the trials since the measure would eventually converge to a maximum as data accumulate.

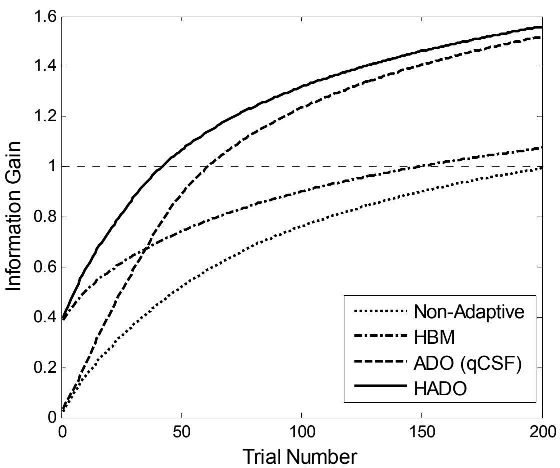

Figure 4: Comparison of information gain of the four experimental design methods for estimation of CSFs.

The HBM condition (dash-dot curve), which employs the hierarchical modeling alone and no stimulus selection technique, enjoys the prior information provided by the hierarchical structure at the start of a session and exhibits greater information gain than the ADO method until it reaches trial 34. However, due to the lack of stimulus optimization, the speed of information gain is considerably slower, taking 152 trials to attain baseline performance. The non-adaptive approach (dotted curve), with neither prior information nor design optimization, shows the lowest level of performance.

Information gain analyzed above may be viewed as a summary statistic, useful for evaluating the measurement methods under comparison. Not surprisingly, we were able to observe the same profile of performance differences in estimating the CSF parameters. Figure 5 shows the comparison of parameter estimation errors for each of the four methods. Error was quantified in terms of root mean squared error (RMSE; y-axis) from the known, underlying parameter value over 200 trials (x-axis). Because we observed the same trend for all parameters, results for the first parameter (peak sensitivity) is shown for simplicity.

As with the case of information gain, HADO benefits from the informative prior through the hierarchical model as well as the optimal stimuli through design optimization, exhibiting the lowest RMSE of all methods' from the start to the end of a session. The benefit of the prior information is also apparent in the HBM condition, making the estimates more accurate than with the uninformed, ADO method for the initial 40 trials, but the advantage is eclipsed in further trials by the effect of design optimization in ADO. In sum, HADO combines the strengths of both ADO and HBM to enjoy both the initial boost in performance and faster decrease in error throughout the experiment.

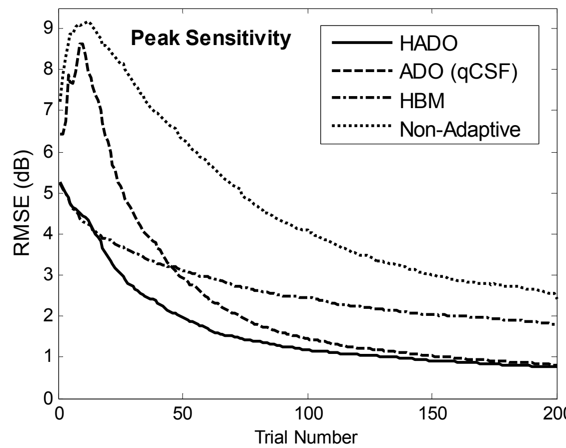

Figure 5: Accuracy of parameter estimation over measurement trials achieved by each of the four measurement methods.

Discussion

The present study demonstrates how hierarchical Bayes modeling (HBM) can be integrated into adaptive design optimization (ADO) to improve the efficiency and accuracy of measurement. The resulting hierarchical adaptive design optimization (HADO) further improves the efficiency of experiments by not only achieving maximal information gain in each experimental trial, but also borrowing information from other experiments. When applied to the problem of estimating a contrast sensitivity function (CSF) in visual psychophysics, HADO achieved an average decrease in parameter estimation error of 38% (from 4.9 dB to 3.1 dB; see Lesmes et al., 2010, for the measurement scale of errors) over conventional ADO, under the scenario that a new session could afford to make only 30 measurement trials.

Although the simulation study served the purpose of demonstrating the benefit of the hierarchical adaptive methodology, the full potential of HADO should be greater than that demonstrated in our particular example. The level of improvement possible with HADO depends on the sophistication of the hierarchical model itself. In our case, the model was based on a simple hypothesis that a newly tested individual belongs to the population from which all other individuals have been drawn. It conveys no further specific information about the likely state of a new individual (e.g., his or her membership to a sub-population is unknown).

There are various situations in which hierarchical modeling can take better advantage of the data-generating structure. For example, although modeled behavioral traits vary across individuals, they may covary with other variables that can be easily observed, such as demographic information (e.g., age, gender, occupation, etc.) or other measurement data (e.g., contrast sensitivity correlates with measures of visual acuity - eye chart test). In this case, a general, multivariate regression or ANOVA model may be employed as the upper-level structure to utilize such auxiliary information to define a more detailed relationship between individuals. This greater detail in the hierarchical model should promote efficient measurement by providing more precise information about the state of future individuals.

Another situation in which hierarchical modeling would be beneficial is when a measurement is made after some treatment and it is sensible or even well known that the follow-up test has a particular direction of change in its outcome (i.e., increase or decrease). Taking this scenario one step further, a battery of tests may be assumed to exhibit profiles that are characteristic of certain groups of individuals. The higher-level structure can also be modeled (e.g., by an autoregressive model) to account for such transitional variability in terms of the parameters of the measurement model. With these kinds of structure built in the hierarchical model, HADO can be used to infer quickly the state of new individuals.

Conclusion

Science and society benefit when data collection is efficient with no loss of accuracy. The proposed HADO framework, which judiciously integrates the best features of design optimization and hierarchical modeling, is an exciting new tool that can significantly improve upon the current state of the art in experimental design, enhancing both measurement and inference. This theoretically well-justified and widely applicable experimental tool should help accelerate the pace of scientific advancement in behavioral and neural sciences.

Acknowledgments

This research is supported by National Institute of Health Grant R01-MH093838 to J.I.M and M.A.P., as well as National Eye Institute Grant R01-EY021553-01 to Z.-L.L. The present conference paper is based largely on Kim et al. (under review) that has been submitted for journal publication.

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.

Atkinson, A., & Donev, A. (1992). *Optimum experimental designs*. Oxford University Press.

Box, G. F. B., & Hill, W. J. (1967). Discrimination among mechanistic models. *Technometrics*, *9*, 57-71.

Cavagnaro, D. R., Pitt, M. A., Gonzalez, R., & Myung, J. I. (2013). Discriminating among probability weighting functions using adaptive design optimization. *Journal of Risk and Uncertainty, 47(3)*, 255-289.

Cavagnaro, D. R., Tang, Y., Myung, J. I. & Pitt, M. A. (2009). Better data with fewer participants and trials: Improving experimental efficiency with adaptive design optimization. In N. A. Taatgen & H. van Rijn (eds.), Austin, TX: Cognitive Science Society.

Comerford, J. P. (1983). Vision evaluation using contrast sensitivity functions. *American Journal of Optometry and Physiological Optics, 60*, 394–398.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. Hoboken, New Jersey: John Wiley & Sons, Inc.

DiMattina, C., & Zhang, K. (2008). How optimal stimuli for sensory neurons are constrained by network architecture. *Neural Computation*, *20*, 668-708.

Good, I. J. (1965). *The estimation of probabilities: An essay on modern Bayesian methods*. Cambridge, MA: MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning.* Springer.

Kim, W., Pitt, M.A., Lu, Z.-L., Steyvers, M. & Myung, J.I. (under review). A hierarchical adaptive approach to optimal experimental design.

Lee, M. D. (2006). A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cognitive Science*, 30, 55-580.

Lesmes, L. A., Jeon, S.-T., Lu, Z.-L., & Dosher, B. A. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick TvC method. *Vision Research, 46*, 3160-3176.

Lesmes, L. A., Lu, Z.-L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quck CSF method. *Journal of Vision, 10,* 1-21.

Lindley, D.V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, *27*(4), 986-1005.

Miller, F., Dette, H., & Guilbaud, O. (2007). Optimal designs for estimating the interesting part of a dose-effect curve. *Journal of Biopharmaceutical Statistics*, *17,* 6.

Myung, J. I., Cavagnaro, D. R., & Pitt, M. A. (2013). A tutorial on adaptive design optimization. *Journal of Mathematical Psychology*, *57*, 53-67.

Tang, Y., Young, C., Myung, J. I., Pitt, M. A., & Opfer, J. (2010). Optimal inference and feedback for representational change. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (p.2572-2577). Austin, TX: Cognitive Science Society.