Hierarchical Bayesian Modeling of Individual Differences in Texture Discrimination

Timothy N. Rubin (trubin@uci.edu)
Michael D. Lee (mdlee@uci.edu)
Charles F. Chubb (cchubb@uci.edu)
Department of Cognitive Sciences, 3151 Social Sciences Plaza A
University of California, Irvine, CA 92697, USA

Abstract

The ability to identify a target texture in a visual display is a basic capability of the human visual system. Traditionally, as with much psychophysical modeling in the cognitive sciences, models of texture discrimination have been fit to individual subject data. By estimating model parameters independently, this approach emphasizes individual differences, and does not model the similarities between subjects. We consider alternative assumptions about individual differences in texture discrimination, using a standard psychometric (or psychophysical) model and previously studied data. In particular, we show how a hierarchical Bayesian approach can capture both the similarities and differences over subjects in a theoretically satisfying way. We also show that the hierarchical Bayesian approach has a number of methodological advantages over existing analyses, including improving parameter estimation when data are either sparse or missing, and improving model predictions when generalizing to new or different stimuli.

Introduction

A key question in understanding any phenomenon in the cognitive sciences relates to individual differences. From basic visual and auditory abilities to more abstract memory, learning and decision-making abilities, there are both similarities and differences between people. A goal of any model should be to provide an account of these similarities and differences.

Often in modeling higher-order cognition, the assumption is made that there are no important individual differences, because the modeling focuses on data averaged or aggregated across subjects (Lee & Webb 2005). In modeling lower-order capabilities, such as in vision and audition, however, the opposite assumption is often made; Models are fit independently to the data of each individual subject, and so the similarities between people are not modeled.

In this paper, we follow current work in modeling individual differences for both low- and high-level capabilities, and adopt hierarchical Bayesian methods to capture both the similarities and individual differences between subjects (e.g., Rouder & Lu 2005, Lee 2008).

Our study of hierarchical Bayesian methods for modeling individual differences takes the form of a case study in a fundamentally important low-level visual capability: the discrimination of texture. Human vision preattentively segments the visual world into different regions based not only on color and light intensity but also on texture. In order to detect a patch of one texture in a background of another in a brief display, human vision must embody one or more fast, spatially parallel mechanisms that are differentially activated by the target versus the background textures.

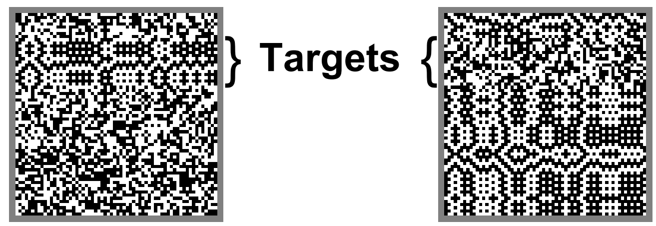

Figure 1: Example of stimuli used by Victor et al. (2005). The left stimulus has a target composed of structured texture against random background. The right stimulus has a target of random texture against a structured texture background.

Texture Discrimination Data

We reconsider data collected by Victor, Chubb and Conte (2005), who analyzed human visual sensitivity to change within a two-dimensional texture space, using an experiment in which subjects were given a four-way forced-choice task and had to locate a target texture against a background. As shown in Figure 1, the stimuli were 64×64 element arrays of black and white squares. Rectangular (either 16×64 or 64×16 element) targets were embedded within the background near one of four stimulus edges. On half of the trials, the target was composed of a structured texture (iid with white and black elements occurring with equal probability). On the other half of the trials, the target was the random texture, and the background was the structured texture.

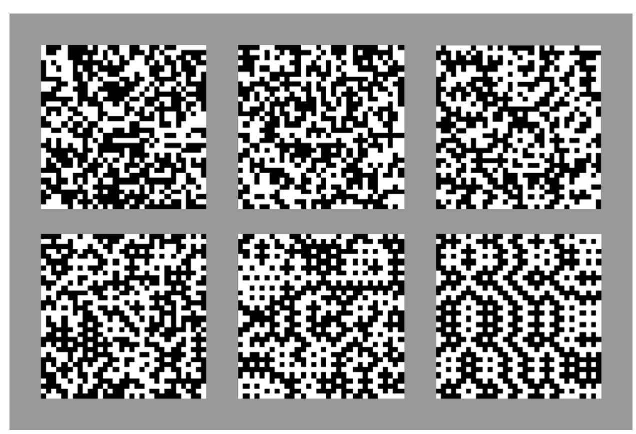

Figure 2: Textures used in the stimuli: Oddnesses 0.5, 0.6, 0.7 (left to right, top row), and 0.8, 0.9. 1.0 (bottom row)

We reanalyze a subset of the data collected by Victor, Chubb and Conte (2005). In this subset, the structured texture varied in "odd-parity" or more simply "oddness": i.e., in the proportion of 2×2 element-blocks that contain odd numbers of whites (and blacks). In particular, the structured textures considered here had oddnesses of 0.6, 0.7, 0.8, 0.9 and 1.0. The random texture (the top-left panel of Figure 2) used in each stimulus had oddness 0.5 (i.e., in the random texture, 2×2 blocks were equally likely to have an even or an odd number of white elements).

Texture Discrimination Modeling

Victor et al. (2005) modeled the effect of texture oddness on performance using a Weibull psychophysical function, so that the probability the *i*th subject successfully discriminated the *j*th oddness, o_j , is given by:

$$\theta_{ij} = 1 - \frac{3}{4}\exp\left[-\left(\frac{o_j}{\lambda_i}\right)^{v_i}\right],$$

where the parameter *v* governs the steepness of the Weibull function, or the speed with which the stimuli become more discriminable, and λ is a scale parameter, that can be thought of as the rate at which subjects approach perfect performance relative to oddness. Because chance performance in the task was 25% correct, the Weibull was restricted to values between 0.25 and 1.0.

Victor et al. (2005) modeled the two target and background relationships described in Figure 1 separately; however, data from both conditions were combined in our analyses. These changes mean that the data we modeled take the form of: counts of the number of correct responses k_{ij} out of *n*=120 trials for the *i*th subject at the *j*th oddness.

Three Models of Individual Differences

In our modeling, we retain the core visual modeling assumptions made by Victor et al. (2005), and continue to model the effect of oddness on performance using a Weibull function. This provides a psychophysical model to map physical stimulus properties to individual subject behavior, which we use to evaluate different assumptions about individual differences.

Specifically, we consider three analyses, corresponding to distinct theoretical perspectives on individual differences. The first analysis involves a "No Individual Differences" model in which there are no underlying differences in subjects' sensitivity to texture oddnesses. The second analysis involves a "Full Individual Differences" model in which each subject's sensitivity function is fully independent of other subjects. The third analysis involves a "Hierarchical Individual Differences" model in which each individual has their own sensitivity function, but its form is constrained by assumptions about distribution of individual differences at the population level.

No Individual Differences (NID) Model

The NID model assumes that there are no individual differences in sensitivity to texture oddness. This means that it infers a single *v* and λ parameter for a single Weibull function that applies to all subjects. According to the NID account, all differences between the performance of subjects for the same stimuli are random noise.

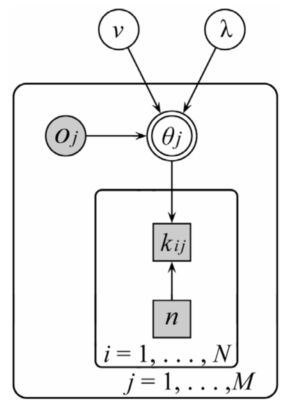

Figure 3: The "No Individual Differences" (NID) model.

Figure 3 shows our implementation of the NID model using the language of probabilistic graphical modeling (see Lee 2008, for a psychologically-oriented introduction). The basic idea is that variables are represented by nodes in a graph, with the graph structure indicating dependencies between nodes, with children depending on their parents. Observed variables (e.g., data) are shaded, while unobserved variables (e.g., underlying psychological parameters) are not shaded. Enclosing plates denote independent replication in the graph structure.

In Figure 3, the probability that k_{ij} out of *n* textures will be discriminated follows a Binomial distribution with probability of success θ_j, and this probability is determined by the Weibull psychophysical model using the *v* and *λ* parameters, and the known texture oddness. The inner plate repeats this modeling over the *i*=1,…,*N* subjects, and the outer plate repeats over the *j*=1,…,*M* oddnesses. The final part of the model gives standard uninformative prior probability distributions to the *v* and *λ* parameters.

Results All of our modeling results are achieved using modern computational Bayesian methods, through sampling from the posterior distribution of the using Markov-Chain Monte-Carlo. We implemented these analyses in WinBUGS (Lunn, Thomas, Best and Speigelhalter, 2000).

Table 1: Posterior parameter means and 2.5% and 97.5% credible intervals for the NID model.

Parameter	Mean	2.5% CI	97.5% CI
v	2.22	1.96	2.51
λ	0.59	0.57	0.62

Table 1 summarizes the marginal posterior distributions for the v and λ parameters, in terms of the mean and 2.5% and 97.5% credible intervals. Figure 4 shows the posterior predictions of the data for each subject, which is a standard Bayesian method for assessing the descriptive adequacy of a model. Because of the assumption that there are no individual differences, the predictions are the same for each subject. We also calculated the arc-sin transformed root mean square deviation (RMSD) as a standard psychophysical summary measure of the goodness-of-fit. For the NID model, the RMSD was 3.83.

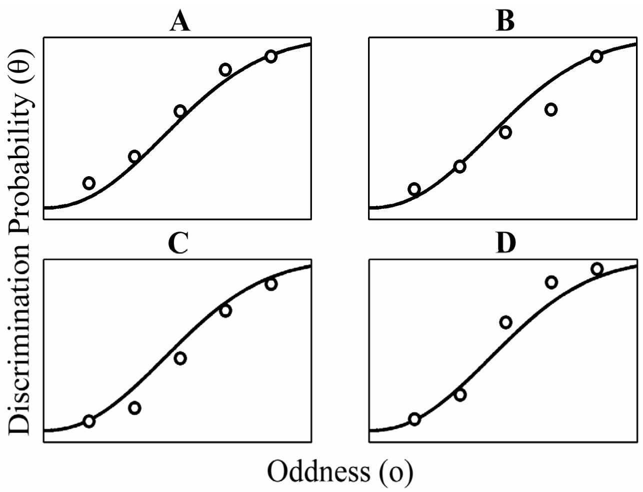

Figure 4: The posterior prediction of the NID model (lines) and the data of each of the four Subjects A-D (circles).

Although Figure 4 suggests the NID model predictions provide a reasonable first-order approximation to the data, it is also clear that the model provides an incomplete description, due to the presence of systematic individual deviations from the model predictions. For example, the model underestimates performance at all oddnesses for Subject C, but overestimates performance at all but one oddness for Subject A.

Full Individual Differences (FID) Model

Our implementation of the FID model is shown in Figure 5.

In this model, differences between the performance of subjects are assumed to result from genuine differences in their abilities. This means each subject has their own parameterization of the Weibull function, determined independently for each subject, that governs their probability of success at each modulation strength. In the graphical model in Figure 5, this key change is made by extending the inner plate (corresponding to subject number) to contain v_i and λ_i parameters for each subject.

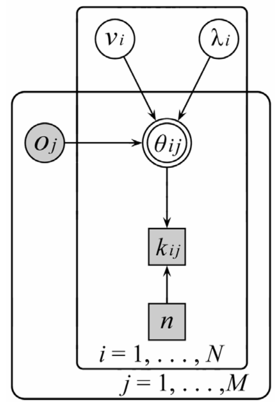

Figure 5: The "Full Individual Differences" (FID) model.

Results Table 2 summarizes the marginal posterior distributions for both parameters for each subject.

Table 2: Posterior parameter means and 2.5% and 97.5% credible intervals for each subject for the FID model.

Parameter	Mean	2.5% CI	97.5% CI
v_A	1.84	1.43	2.33
λ_A	0.55	0.50	0.60
v_B	2.09	1.50	2.77
λ_B	0.65	0.60	0.71
v_C	2.51	1.91	3.19
λ_C	0.65	0.61	0.70
v_D	2.47	1.97	3.00
λ_D	0.53	0.49	0.57

Figure 6 shows the posterior predictions, which now capture the variation between individual subjects. The RMSD of the FID model was 2.01, indicating a large improvement over the NID model.

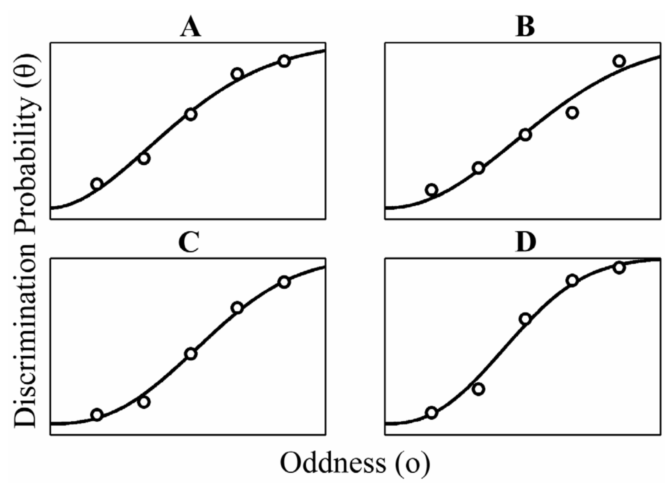

Figure 6: The posterior prediction of the FID model (lines) and the data of each of the four Subjects A-D (circles).

Hierarchical Individual Differences (HID) Model

Figure 7 shows the HID model, which assumes that there are individual differences in the v and λ parameters, but that these differences have structure across the subjects, rather than being free to vary independently. The structure we assume is that both the v and λ parameters are draws from Gaussian distributions, and so have population-level means (μ_v and μ_λ) and variances (σ_v and σ_λ).

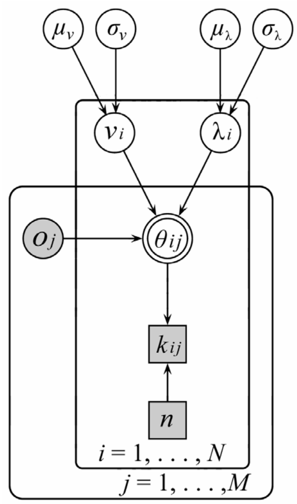

Figure 7: The "Hierarchical Individual Differences" (HID) model.

Inference for this hierarchical model involves both the parameters of the Gaussian distributions and the specific v and λ parameters for each subject. Because v and λ are now generated from the Gaussian distributions, they are not assigned priors. Rather, we now assign standard vague priors on the μ_v, μ_λ, σ_v and σ_λ parameters of the Gaussians.

Results Table 3 summarizes the marginal posterior distributions for both parameters for each subject

Table 3: Posterior parameter means and 2.5% and 97.5% credible intervals for each subject under the HID model.

Parameter	Mean	2.5% CI	97.5% CI
v_A	2.13	1.66	2.53
λ_A	0.56	0.52	0.60
v_B	2.23	1.80	2.67
λ_B	0.64	0.60	0.69
v_C	2.36	1.97	2.89
λ_C	0.64	0.60	0.69
v_D	2.37	2.00	2.84
λ_D	0.53	0.50	0.57

Figure 8 shows the posterior predictions of the HID model, overlaid on the already presented NID and FID predictions from Figures 4 and 6. It is clear that the HID model makes extremely similar predictions to the FID model. Figure 8 also illustrates the impact of assuming structured individual differences; the predictions for each subject are pulled towards the group mean, because the Weibull parameters are modeled as draws from an over-arching Normal distribution.

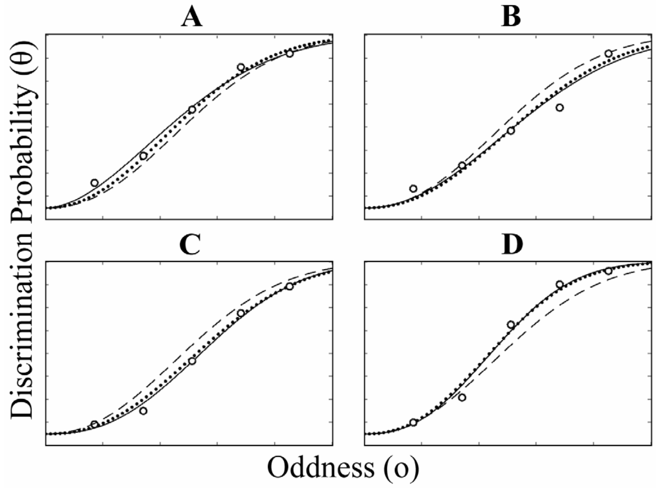

Figure 8: The posterior predictions of the NID, FID and HID models (dashed, solid and dotted lines, respectively) against the data of each of the four Subjects A-D (circles).

The RMSD for the HID model is 2.15, slightly worse than the 2.01 for the FID model. Given they make similar predictions, and the FID model fares quantitatively better in terms of data fit, it is reasonable to ask what, if any, advantages, are conferred by the HID model, and the general approach of using structured individual differences. The remainder of our results address this challenge.

Advantages of the Hierarchical Model

In addition to describing existing data accurately, a good model should be able to make predictions about unknown data. Indeed, being able to generalize successfully to new and different situations is one of the key tests of a model, and provides strong grounds for believing a model is useful. To address this issue, we tested the ability of the FID and HID models to predict the probability of discrimination for individual subjects at oddnesses for which data were withheld, and to predict individual sensitivity functions using scarce or missing data for one or more subjects.

Generalizing to New Oddnesses

We constructed four specific tests by removing for all subjects: oddness 3; oddnesses 1 and 5; oddnesses 2 and 4; and oddnesses 2, 3 and 4. RMSDs for the model predictions are shown in Table 4. Overall, the HID Model performed better than the FID model in predicting individual performance at missing oddnesses.

The relatively poor performance of the FID Model is a direct consequence of the assumptions it makes about individual differences. When data points are missing, the model attempts inference based on only the remaining data points for that subject. If these remaining data points are not representative of the overall psychophysical function for the subject, the inferences and subsequent predictions will be inaccurate. The HID Model, in contrast, is able to use data from other subjects to help make inferences at the missing oddnesses.

Table 4: RMSDs of FID and HID model predictions at missing oddnesses

Removed Oddnesses	FID	HID
3	1.66	1.56
1, 5	5.13	3.03
2, 4	3.71	3.81
2, 3, 4	4.11	4.38
Mean	3.65	3.19

Figure 9 gives an example of the important difference between the FID and HID models when generalizing to stimuli for which data were not available. It shows the posterior predictions of the two models for Subject B in the second condition. Because the subject had near-linear data for the three middle oddnesses strengths, the FID Model generates a Weibull function that does not resemble any of the functions found when all data were modeled, and hence makes a drastic mis-prediction for the fifth oddness. The HID Model on the other hand uses the data of the other three subjects to guide predictions about Subject B at extreme modulations, which results in a better predicted psychophysical function.

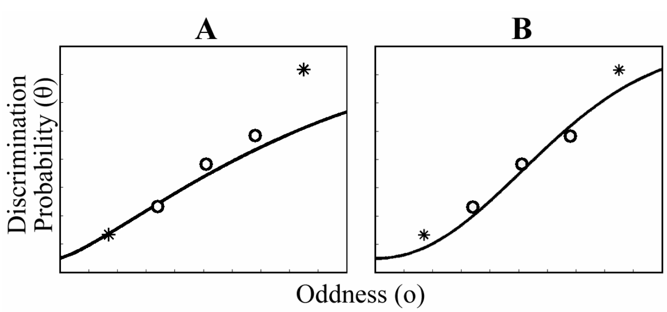

Figure 9: Posterior predictions for the FID (panel A) and HID (panel B) models for the data of Subject B when oddnesses 1 and 5 are missing (denoted by stars).

Predictions from Scarce or Missing Data

To study the effect of scarce or missing data for one or all subjects, the models were applied to data sets impoverished in different ways. Scarce data were created by dividing the number of correct trials and total trials at each data point by 10, 20, 30 or 60, and then rounding to the nearest integer. Data thus took the form of a correct count out of 12, 6, 4 or 2 trials, respectively. Although this rounding procedure is not equivalent to a true reduced trial experiment, it provides an easy and useful approximation to one.

Scarce Data for Single Subjects For these analyses, the available information was modified so that each subject, in turn, had either scarce data (i.e., truncated to 6 trials) or their data were missing entirely. In each analysis, the data for the other subjects were left intact. Table 5 shows the RMSDs of the FID and HID model predictions for each subject under both the scarce and missing data analyses.

Table 5: RMSDs for each subject, and the mean for all subjects, under the FID and HID models, in the scarce and missing data analyses.

	Scarce Data		Missing Data	
Subject	FID	HID	FID	HID
A	4.55	2.46	9.63	3.15
B	6.56	4.20	9.84	4.72
C	7.54	2.49	11.45	4.60
D	7.40	5.07	13.32	6.22
Mean	6.51	3.55	11.06	4.67

Across all conditions, the HID Model significantly outperformed the FID Model. The relative resilience of the HID Model to scarce data is again a consequence of its assumptions about the structure of individual differences. In both types of conditions, the model infers normal distributions using all available data. Since the other subjects provide a large number of data, the model infers group means of the parameters reasonably accurately. In the scarce data conditions, the model then uses the limited available data for a subject to shift the individual parameters away from the group mean. In the missing data conditions, the HID model resembles the NID Model, in the sense that it infers the data of missing subjects data based on the group mean. The differences in the performance of the HID model across the missing-subject conditions are due the relative deviation of the specific subject's data from the group mean.

The FID model, in contrast, infers the parameters for each subject based only on individual subject data. The result in the scarce data condition is that a Weibull is fit as closely as possible to noisy data. The large variations in the accuracy of model predictions across subjects are due to the varying degrees to which the 6 trial samples are representative of the full data set. In the missing data condition, the only information the FID model can draw on is the priors, leading to very poor prediction.

Scarce Data for All Subjects Our final analyses examined prediction accuracy of the models when data were scarce for all subjects. The RMSDs in four truncation conditions, corresponding to 12, 6, 4 and 2 trials across all subjects are presented in Table 6. Posterior predictions for Subject C (which are representative of the other subjects) are shown in Figure 11 for both models over an additional set of truncations, involving 120, 40 and 6 trials. In both Table 6 and Figure 10, the HID model proves to be significantly more accurate than the FID model, further demonstrating its resilience when limited data are available. When the HID model does not have access to a large number of trials for any individual subject, it is able to make inferences about the group distribution by combining the data from all four subjects. It then automatically adjusts the posterior predictions away from the group mean for individual subjects based on only their data. The FID model, by contrast, uses the data from each subject in isolation. As the number of trials decreases, the information available to infer individual subject parameters quickly decreases. This leads to less certainty and accuracy in the posterior distributions of the parameters, and the greater variability in posterior prediction, especially evident in Figure 10.

Table 6: RMSDs for the FID and HID models at different levels of truncation.

Truncation	FID	HID
12	4.27	2.65
6	6.71	3.62
4	6.50	2.77
2	9.81	7.26
Mean	6.82	4.08

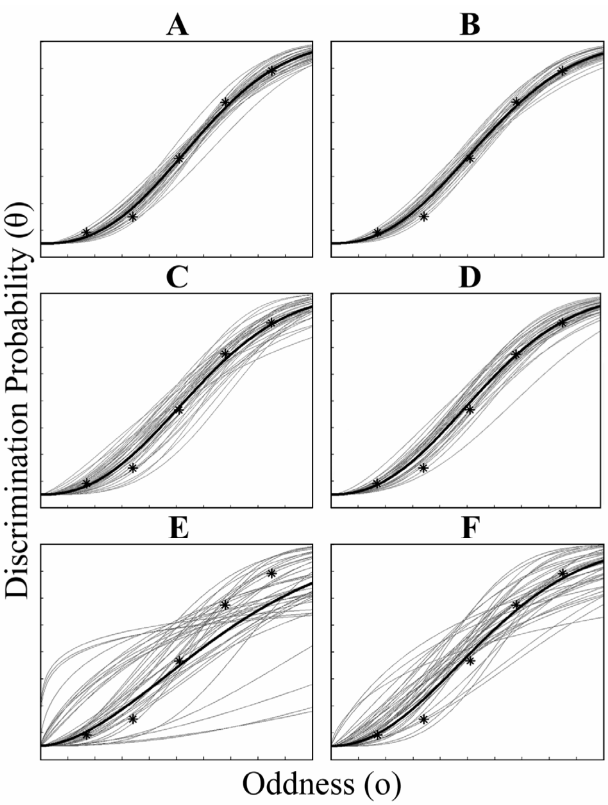

Figure 10: Predictions for Subject C of the FID model (A, C, E in the left column) and the HID Model (B, D, F in the right column) for sparse data. The different analyses involve using 120 (A, B) 40 (C, D) and 6 (E, F) trials. The predictive lines are the psychometric functions generated by 40 random samples from the joint posterior parameter distributions.

Conclusions

Modeling of low-level phenomena in the cognitive sciences typically makes the assumption that there are no important similarities between individuals, and focuses on fitting models to individual subject data. One of the primary motivations for this approach is that, quantitatively, it generates models that best account for the observed data. However, on a theoretical level, the modeling approach runs counter to the general assumption in psychology that individual differences follow some orderly population-level distribution, as can be captured by hierarchical models.

In this study, we considered three modeling approaches, corresponding to different theoretical accounts of individual differences. The NID model proved inadequate due to its failure to describe the systematic individual variation in the observed data. By contrast, both of the FID and HID models incorporated individual differences, and generated posterior predictions that described the original data well.

We examined whether the FID and HID models could predict the full data set when trained using scarce or missing data for one or more subjects. Across nearly all of these conditions, the HID model was superior. It proved resilient under all but the most extreme case in which data for all subjects was truncated to 2 trials, whereas the FID model was highly sensitive to all forms of data degradation

In the most extreme case of scarce individual data, the theoretical distinction between the models is clearest. For a subject who has no data, the FID model depends entirely on the priors, because it does not model population-level relationships between individuals. In contrast, the relationship between individuals is built into the HID model, allowing it to make reasonable inferences about even for subjects who have yet to provide data. Finally, and perhaps most importantly to psychophysical experimentation, the HID model generalized subject data to untested oddnesses better than the FID model.

Overall, our results suggest that, for both practical and theoretical reasons, hierarchical Bayesian models provide an appealing approach to understanding and incorporating individual differences in psychophysical modeling.

References

Lee, M.D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psych. Bull. & Rev., 15,* 1-15.

Lee, M.D., & Webb, M.R. (2005). Modeling individual differences in cognition. *Psych. Bull. & Rev.*, *12*, 605-621.

Lunn, D.J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325-337.

Rouder, J.N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psych. Bull. & Rev., 32*, 573–604.

Victor, J.D., Chubb, C., & Conte, M.M. (2005). Interaction of luminance and higher-order statistics in texture discrimination. *Vision Research, 45*, 311-328.