

Hierarchical Bayesian modeling of contrast sensitivity functions in a within-subject design

Yukai Zhao

Center for Neural Science, New York University,
New York, NY, USA



Luis Andres Lesmes

Adaptive Sensory Technology Inc., San Diego, CA, USA



Fang Hou

School of Ophthalmology & Optometry and Eye Hospital,
Wenzhou Medical University, Wenzhou, Zhejiang, China



Division of Arts and Sciences, NYU Shanghai,
Shanghai, China

Center for Neural Science and Department of Psychology,
New York University, New York, NY, USA

NYU-ECNU Institute of Brain and Cognitive Science,
Shanghai, China



Zhong-Lin Lu

Recent development of the quick contrast sensitivity function (qCSF) method has made it possible to obtain accurate, precise, and efficient contrast sensitivity function (CSF) assessment. To improve statistical inference on CSF changes in a within-subject design, we developed a hierarchical Bayesian model (HBM) to compute the joint distribution of CSF parameters and hyperparameters at test, subject, and population levels, utilizing information within- and between-subjects and experimental conditions. We evaluated the performance of the HBM relative to a non-hierarchical Bayesian inference procedure (BIP) on an existing CSF dataset of 112 subjects obtained with the qCSF method in three luminance conditions (Hou, Lesmes, Kim, Gu, Pitt, Myung, & Lu, 2016). We found that the average d' s of the area under log CSF (AULCSF) and CSF parameters between pairs of luminance conditions at the test-level from the HBM were 33.5% and 103.3% greater than those from the BIP analysis of AULCSF. The increased d' resulted in greater statistical differences between experimental conditions across subjects. In addition, simulations showed that the HBM generated accurate and precise CSF parameter estimates. These results have strong implications for the application of HBM in clinical trials and patient care.

filtered stimuli over a wide range of spatial frequencies, provides a comprehensive measure of spatial vision (Ginsburg, 1981; Ginsburg, 2003; Hess, 1981). It is closely related to daily visual functions (Ginsburg, 2003), and can better quantify deficits in spatial vision than visual acuity (Jindra & Zemon, 1989; Marmor, 1986). It has long been recognized that the CSF provides important information for monitoring progression of vision change and evaluating treatment efficacy in eye diseases (Bellucci, Scialdone, Buratto, Morselli, Chiarego, Criscuolo, Criscuoli, Moretti, & Piers, 2005; Ginsburg, 2006; Loshin & White, 1984; Levi & Li, 2009; Tan & Fong, 2008; Zhou, Huang, Xu, Tao, Qiu, Li, & Lu, 2006).

Despite its clinical promise, precise and efficient CSF assessment has presented a challenge. The CSF charts provide a fast but imprecise assessment of contrast sensitivity due to coarse sampling of both spatial frequency and stimulus contrast (Bradley, Hook, & Haeseker, 1991; Buhren, Terzi, Bach, Wesemann, & Kohnen, 2006; Hohberger, Laemmer, Adler, Juenemann, & Horn, 2007; Pesudovs, Hazel, Doran, & Elliott, 2004; van Gaalen, Jansonius, Koopmans, Terwee, & Kooijman, 2009). On the other hand, the long testing time (30–60 minutes) required for measuring the CSF with conventional psychophysical methods has prevented their clinical applications (Kelly & Savoie, 1973; Treutwein, 1995). The quick contrast sensitivity function (qCSF) was developed to address the challenges (Lesmes, Lu, Baek, & Albright, 2010). Based on active learning principles, it estimates

Introduction

The contrast sensitivity function (CSF), which quantifies the visibility ($1/\text{threshold}$) of narrow-band

Citation: Zhao, Y., Lesmes, L. A., Hou, F., & Lu, Z.-L. (2021). Hierarchical Bayesian modeling of contrast sensitivity functions in a within-subject design. *Journal of Vision*, 21(12):9, 1–17, <https://doi.org/10.1167/jov.21.12.9>.



the parameters of the CSF in a Bayesian adaptive framework (Kontsevich & Tyler, 1999; Lu & Dosher, 2013; Watson, 2017; Watson & Pelli, 1983). A recent qCSF implementation with a 10-letter identification task enabled assessment of the CSF with a 0.10 log unit standard deviation in about 20 trials (approximately 2 minutes) and reduced the standard deviation of the estimates by 50% (Hou, Lesmes, Bex, Dorr, & Lu, 2015). Accurate and precise qCSF estimates have been obtained in both normal (Reynaud, Tang, Zhou, & Hess, 2014; Rosén, Lundström, Venkataraman, Winter, & Unsbo, 2014) and clinical populations (Hou, Huang, Lesmes, Feng, Tao, Zhou, & Lu, 2010; Jia, Zhou, Lu, Lesmes, & Huang, 2015; Joltikov, de Castro, Davila, Anand, Khan, Farbman, Jackson, Johnson, & Gardner, 2017; Lesmes, Jackson & Bex, 2013; Lesmes, Wallis, Jackson, & Bex, 2013; Lesmes, Wallis, Lu, Jackson, & Bex, 2012; Lin, Mihailovic, West, Johnson, Friedman, Kong, & Ramulu, 2018; Ou, Lesmes, Christie, Denlar, & Csaky, 2021; Ramulu, Dave, & Friedman, 2015; Rosen, Jayaraj, Bharadwaj, Weeber, Van der Mooren, & Piers, 2015; Stellmann, Young, Pottgen, Dorr, & Heesen, 2015; Thomas, Silverman, Vingopoulos, Kasetty, Yu, Kim, Omari, Joltikov, Choi, Kim, Zacks, & Miller, 2021; Vingopoulos, Wai, Katz, Vavvas, Kim, & Miller, 2021; Wai, Vingopoulos, Garg, Kasetty, Silverman, Katz, Láins, Miller, Husain, Vavvas, Kim, & Miller, 2021; Yan, Hou, Lu, Hu, & Huang, 2017).

A Bayesian Inference Procedure (BIP) has been developed to make statistical inference on CSF changes in a within-subject design based on CSF metrics extracted from each subject in each experimental condition (Hou, et al., 2016; Kuss, Jäkel, & Wichmann, 2005; Prins, 2013; Schütt, Harmeling, Macke, & Wichmann, 2016). Because it scores each test independently with an uninformative prior without considering potential relationships of CSF parameters across subjects and experimental conditions, the BIP may have overestimated the variance of each test and resulted in reduced statistical power (Borm, Fransen, & Lemmens, 2007; Egbewale, Lewis, & Sim, 2014; Wilcox, 2012). In addition, a single summary metric, the area under log CSF (AULCSF), is usually used to compare CSFs in different experimental conditions, potentially leaving out information in the multidimensional joint distribution of CSF parameters.

In this study, we developed a Hierarchical Bayesian Model (HBM) to reduce the variability of estimated CSF parameters for each test and to further improve the ability to detect between-condition CSF changes in a within-subject design. The HBM is a generative model framework that uses Bayes' rule to quantify the joint distribution of test-, subject-, and population-level parameters and hyperparameters (Kruschke, 2015; Lee, 2006; Lee,

2011; Rouder & Lu, 2005; Wilson, Cranmer, & Lu, 2020). It explicitly quantifies the covariance of the hyperparameters and parameters (Daniels & Kass, 1999; Klotzke & Fox, 2019; Thall, Wathen, Bekele, Champlin, Baker, & Benjamin, 2003; Wang, Lin, & Nelson, 2020; Yang, Zhu, Choi, & Cox, 2016). By sharing information within and across levels via conditional dependencies, it reduces the variance of the test-level estimates through (1) decomposition of variabilities from different sources (test, subject, and population) with parameters and hyperparameters (Song, Behmanesh, Moaveni, & Papadimitriou, 2020), and (2) shrinkage of the estimated parameters at the lower levels toward the mean of the higher levels when there is not sufficient data at the lower level (Kruschke, 2015; Rouder & Lu, 2005; Rouder, Sun, Speckman, Lu, & Zhou, 2003).

Although it has been used in many different disciplines, such as astronomy (Thrane & Talbot, 2019), ecology (Reum, Hovel, & Greene, 2015; Wikle, 2003), genetics (Storz & Beaumont, 2002), machine learning (Li & Perona, 2005), cognitive science (Ahn, Krawitz, Kim, Busmeyer, & Brown, 2011; Lee, 2006; Lee & Mumford, 2003; Merkle, Smithson, & Verkuilen, 2011; Molloy, Bahg, Li, Steyvers, Lu, & Turner, 2018; Molloy, Bahg, Lu, & Turner, 2019; Rouder & Lu, 2005; Rouder et al., 2003; Wilson et al., 2020) and visual acuity (Zhao, Lesmes, Dorr, & Lu, 2021), HBM has not been applied to analyze the CSF. Here, we develop a three-level HBM to model the entire CSF dataset in a single-factor (luminance), multi-condition (3 luminance conditions), and within-subject experiment design. We modeled the data with CSF parameters at the test level and hyperparameters at the individual and population levels, with conditional dependencies across levels. We evaluated the performance of the HBM relative to the BIP using an existing dataset of 112 subjects tested with qCSF in three luminance conditions (Hou et al., 2016), which was collected to mimic mild, medium, and large CSF changes observed in clinical settings (Bellmann, Unnebrink, Rubin, Miller, & Holz, 2003; Haymes, Roberts, Cruess, Nicolela, LeBlanc, Ramsey, Chauhan, & Artes, 2006; Kalia, Lesmes, Dorr, Gandhi, Chatterjee, Ganesh, Bex, & Sinha, 2014; Kleiner, Enger, Alexander, & Fine, 1988; Midena, Degli Angeli, Blarzino, Valenti, & Segato, 1997; Owsley, Sekuler, & Siemsen, 1983). In addition, a simulation study was conducted to evaluate and compare the accuracy and precision of the estimates from the HBM and BIP. We hypothesized that, relative to the BIP, the HBM would reduce the variability of the estimated CSF parameters from each test, increase the d 's of CSF changes between luminance conditions for each subject, and improve statistical inference across subjects.

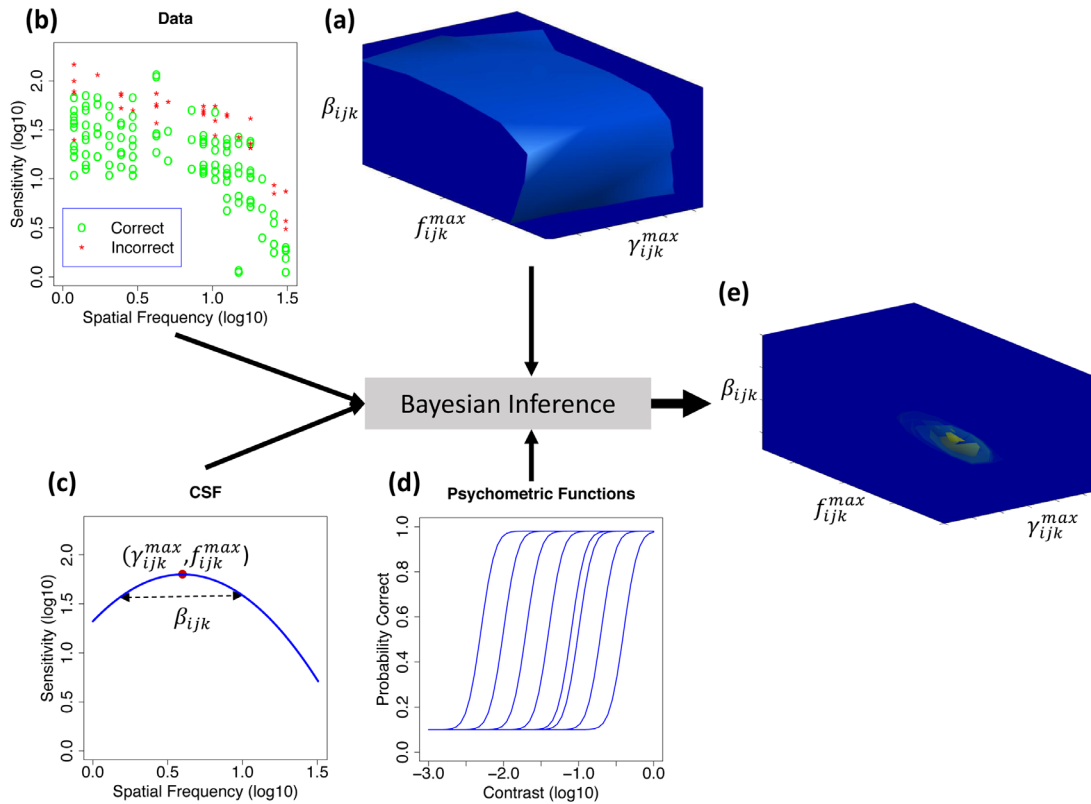


Figure 1. The Bayesian inference procedure (BIP) for a single test. (a) A three-dimensional prior distribution of the CSF parameters. (b) Trial-by-trial data. (c) A CSF model with three parameters. (d) Psychometric functions at different spatial frequencies. (e) A three-dimensional posterior distribution of the CSF parameters.

Bayesian modeling of the CSF

Overview

In a typical within-subject design CSF experiment with multiple conditions, the trial-by-trial data can be organized as $y_{ijkm} = (f_{ijkm}, c_{ijkm}, r_{ijkm})$, where r_{ijkm} , either correct or incorrect, is individual i 's response in trial m of test k in experimental condition j tested with a stimulus of spatial frequency f_{ijkm} and contrast c_{ijkm} . The BIP consists of four components (Hou, et al., 2016): (1) a log-parabola model of the contrast sensitivity function with several parameters, (2) a likelihood function that specifies the probability of making a correct or incorrect response in each stimulus condition, (3) a Bayesian procedure to infer the posterior distribution of the CSF parameters for each subject in each test, and (4) inference based on statistics computed from posterior distributions either at the subject level or aggregated across subjects. In this section, we first provide a brief review of the BIP, and then introduce the HBM.

The Bayesian inference procedure

In the BIP (Figures 1, 2a), the contrast sensitivity $S(f_{ijkm}, \theta_{ijk})$ at spatial frequency f_{ijkm} is modeled with a log parabola function with three parameters, $\theta_{ijk} = (\gamma_{ijk}^{max}, f_{ijk}^{max}, \beta_{ijk})$ (Lesmes et al., 2010; Rohaly & Owsley, 1993; Watson, & Ahumada Jr, 2005):¹

$$\log_{10}(S(f_{ijkm}, \theta_{ijk})) = \log_{10}(\gamma_{ijk}^{max}) - \frac{4}{\log_{10}(2)} \left(\frac{\log_{10}(f_{ijkm}) - \log_{10}(f_{ijk}^{max})}{\beta_{ijk}} \right)^2, \quad (1)$$

where γ_{ijk}^{max} is the peak sensitivity, f_{ijk}^{max} is the peak spatial frequency (cycles/degree), and β_{ijk} is the bandwidth (octaves) at half of the peak sensitivity. The probability of making a correct response is described with a psychometric function (Hou et al., 2015):

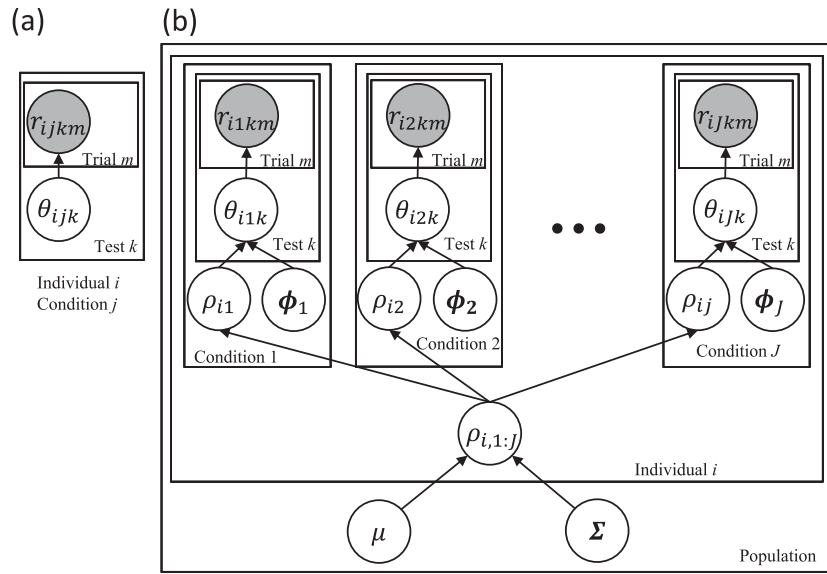


Figure 2. (a) The Bayesian Inference Procedure (BIP) computes the posterior distribution of CSF parameters for each test independently. (b) A three-level hierarchical Bayesian model (HBM) of CSFs across multiple individuals, conditions and tests. At the population level, μ and Σ are the mean and covariance hyperparameters of the population. At the individual level ρ_{ij} and ϕ_j are the mean and covariance hyperparameters of individual i in experimental condition j . At the test level, θ_{ijk} is the CSF parameter of individual i in test k of condition j .

$$p(r_{ijkm} = 1 | \theta_{ijk}, f_{ijkm}, c_{ijkm}) = g + \left(1 - g - \frac{\lambda}{2}\right) \Phi \times \left(\frac{\log_{10}(c_{ijkm}) + \log_{10}(S(f_{ijkm}, \theta_{ijk}))}{\sigma}\right), \quad (2)$$

where g is the guessing rate, λ , usually set to 0.04 (Lesmes et al., 2010; Wichmann & Hill, 2001), is the lapse rate, Φ is the standard cumulative Gaussian function, and σ determines the steepness of the psychometric function. The probability of making an incorrect response is:

$$p(r_{ijkm} = 0 | \theta_{ijk}, f_{ijkm}, c_{ijkm}) = 1 - p(r_{ijkm} = 1 | \theta_{ijk}, f_{ijkm}, c_{ijkm}). \quad (3)$$

Equations 2 and 3 define the likelihood function, that is, the probability of making a correct or incorrect response given the stimulus and CSF parameters in a trial. The goal in most experiments is to infer the CSF parameters from the experimental data, that is, estimate the posterior distribution $p(\theta_{ijk} | Y_{ijk})$ — the distribution of the CSF parameters θ_{ijk} given the experimental data $Y_{ijk} = \{y_{ijkm}\}$, for $m=1, \dots, M$, where M is the total number of trials in a test. This can be accomplished using Bayes' rule:

$$p(\theta_{ijk} | Y_{ijk}) = \frac{\prod_{m=1}^M p(r_{ijkm} | \theta_{ijk}, f_{ijkm}, c_{ijkm}) p_0(\theta_{ijk})}{\int \prod_{m=1}^M p(r_{ijkm} | \theta_{ijk}, f_{ijkm}, c_{ijkm}) p_0(\theta_{ijk}) d\theta_{ijk}}, \quad (4)$$

where $p_0(\theta_{ijk})$ is the prior probability distribution of the CSF parameters for individual i in test k of experimental condition j , which is usually uninformative and the same for all subjects and experimental conditions, and the denominator is the integral across all possible values of θ_{ijk} , and is a constant for a given dataset and BIP.

The hierarchical Bayesian model

We developed a three-level HBM to account for the entire dataset, incorporating conditional dependencies across test, individual, and population levels to improve estimates for each test (see Figure 2b). The HBM is based on three properties: (1) CSF parameters at the test level are conditionally dependent on hyperparameters at the individual level, (2) CSF hyperparameters at the individual level are conditionally dependent on those at the population level (“conditional dependency”), and (3) the probability $p(r_{ijkm} | \theta_{ijk}, f_{ijkm}, c_{ijkm})$ of response r_{ijkm} is determined only by the CSF parameters θ_{ijk} in that test (Equations 2, 3).

In the HBM, the joint distribution of CSF hyperparameter η across all the J experimental conditions at the population level, $p(\eta)$, is modeled as a mixture of $3 \times J$ -dimensional Gaussian distributions \mathcal{N} with mean μ and covariance Σ , which have distributions $p(\mu)$ and $p(\Sigma)$:

$$p(\eta) = \mathcal{N}(\eta, \mu, \Sigma) p(\mu) p(\Sigma). \quad (5)$$

The joint distribution of CSF hyperparameter $\tau_{i,1:J}$ of individual i across all experimental conditions $1:J$ at the individual level, $p(\tau_{i,1:J}|\eta)$, is modeled as mixtures of three-dimensional Gaussian distributions with mean ρ_{ij} and covariance ϕ_j , which have distributions $p(\rho_{i,1:J}|\eta)$ and $p(\phi_j)$:

$$p(\tau_{i,1:J}|\eta) = p(\rho_{i,1:J}|\eta) \prod_{j=1}^J \mathcal{N}(\tau_{ij}, \rho_{ij}, \phi_j) p(\phi_j), \quad (6)$$

where $p(\rho_{i,1:J}|\eta)$ denotes that $\rho_{i,1:J}$ is conditioned on η , and ϕ_j is a 3×3 covariance matrix in experimental condition j . Finally, at the test level, $p(\theta_{ijk}|\tau_{ij})$, the joint distribution of the CSF parameters, θ_{ijk} , is conditioned on τ_{ij} .

The probability of obtaining the entire dataset is computed by probability multiplication:

$$\begin{aligned} & p(Y_{1:I,1:J,1:K,1:M}|X) \\ &= \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{m=1}^M p(r_{ijkm}|\theta_{ijk}, f_{ijkm}, c_{ijkm}) \\ & \quad \times p(\theta_{ijk}|\tau_{ij}) p(\tau_{i,1:J}|\eta) p(\eta) \\ &= \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{m=1}^M p(r_{ijkm}|\theta_{ijk}, f_{ijkm}, c_{ijkm}) \\ & \quad \times p(\theta_{ijk}|\tau_{ij}) \mathcal{N}(\tau_{ij}, \rho_{ij}, \phi_j) p(\phi_j) p(\rho_{i,1:J}|\eta) \\ & \quad \times \mathcal{N}(\eta, \mu, \Sigma) p(\mu) p(\Sigma), \end{aligned} \quad (7)$$

where $X = (\theta_{1:I,1:J,1:K}, \rho_{1:I,1:J}, \phi_{1:J}, \mu, \Sigma)$ are all the parameters and hyperparameters in the HBM.

We can use Bayes' rule to compute the joint posterior distribution of X (Kruschke, 2015; Lee, 2006; Lee, 2011; Rouder & Lu, 2005; Wilson et al., 2020):

$$\begin{aligned} & p(X|Y_{1:I,1:J,1:K,1:M}) \\ &= \frac{\prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{m=1}^M p(r_{ijkm}|\theta_{ijk}, f_{ijkm}, c_{ijkm}) \\ & \quad \times p(\theta_{ijk}|\tau_{ij}) \mathcal{N}(\tau_{ij}, \rho_{ij}, \phi_j) p_0(\phi_j) \\ & \quad \times p(\rho_{i,1:J}|\eta) \mathcal{N}(\eta, \mu, \Sigma) p_0(\mu) p_0(\Sigma)}{\int \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{m=1}^M p(r_{ijkm}|\theta_{ijk}, f_{ijkm}, c_{ijkm}) \\ & \quad \times p(\theta_{ijk}|\tau_{ij}) \mathcal{N}(\tau_{ij}, \rho_{ij}, \phi_j) p_0(\phi_j) p(\rho_{i,1:J}|\eta) \\ & \quad \times \mathcal{N}(\eta, \mu, \Sigma) p_0(\mu) p_0(\Sigma) dX}, \end{aligned} \quad (8)$$

where the denominator is the integral across all possible values of X and is a constant for a given dataset and HBM; $p_0(\mu)$, $p_0(\Sigma)$, and $p_0(\phi_j)$ are the prior distributions.

Methods

Data

The dataset used in this study included 112 college-aged subjects, each tested once ($K = 1$) in three luminance conditions (low = 2.62 cd/m², medium = 20.4 cd/m², and high = 95.4 cd/m²) with the qCSF method (Hou et al., 2016). Each test consisted of 150 trials. Three test trials were presented in each display consisting of three filtered letters of the same size, randomly sampled with replacement from 10 SLOAN letters (C, D, H, K, N, O, R, S, V, and Z), with the center spatial frequency and contrasts of the letters determined by qCSF. Subjects were asked to verbally report the identity of the letters on the screen.

Apparatus

All analysis was conducted on a Dell computer with Intel Xeon W-2145 @ 3.70 GHz CPU (8 cores and 16 threads) and 64 GB installed memory (RAM). The BIP was implemented in Matlab R201Xa (MathWorks Corp., Natick, MA, USA) and the HBM was implemented in JAGS (Plummer, 2003) in R (R Core Team, 2020).

Implementation of the BIP

Because a 10-alternative forced-choice identification task was used in the experiment, we set g to 0.1, and σ to 0.1485 in Equation 2 (Foley & Legge, 1981; Hou et al., 2015; Legge, Kersten, & Burgess, 1987; Lesmes et al., 2010; Lu & Doshier, 1999). Following the qCSF procedure (Hou et al., 2015; Lesmes et al., 2010), we defined a three-dimensional CSF parameter space with 60 log-linearly spaced γ_{ijk}^{max} values between 1.05 and 1050, 40 log-linearly spaced f_{ijk}^{max} values between 0.1 and 20 cycles/degree, and 27 log-linearly spaced β_{ijk} values between 1 and 9 octaves. The weakly informative prior, $p_0(\theta_{ijl})$, identical across all the tests, subjects, and experimental conditions, was defined by a hyperbolic secant function (Lesmes et al., 2010):

$$\begin{aligned} & p_0(\theta_{ijl}) \\ &= \prod_{a=1}^3 \text{sech}(\theta_{a, \text{confidence}} \times (\log_{10}(\theta_a) - \log_{10}(\theta_{a, \text{mode}}))), \end{aligned} \quad (9)$$

where $\text{sech}(x) = \frac{2}{e^x + e^{-x}}$, $\theta_a = \gamma_{ijk}^{max}$, f_{ijk}^{max} , and β_{ijk} for $a = 1, 2$, and 3 , respectively, $\theta_{a, \text{confidence}} = (0.5, 0.5, 0.5)$, and $\theta_{a, \text{mode}} = (100, 1, 3)$.

	L			M			H		
	γ^{max}	f^{max}	β	γ^{max}	f^{max}	β	γ^{max}	f^{max}	β
$\mu_{0,min}$	1.05	0.1	1	1.05	0.1	1	1.05	0.1	1
$\mu_{0,max}$	1050	20	9	1050	20	9	1050	20	9

Table 1. $\mu_{0,min}$ and $\mu_{0,max}$ of the uniform prior of μ . H, high; L, low; M, medium.

The posterior distributions of the CSF parameters $p(\theta_{ij1}|Y_{ij1})$ was computed using Equation 4. Convergence of the BIP solutions was quantified by the half-width of 68.2% credible interval (HWCI: Clayton & Hills, 1993; Edwards, Lindman, & Savage, 1963), equivalent to the standard deviation of the distribution if it is normal. With sufficient number of trials in the qCSF, the HWCI can reach its asymptotic minimum (Hou et al., 2015; Lesmes et al., 2010).

Implementation of the HBM

In the current implementation of the HBM, the prior of μ , $p_0(\mu)$, was a nine-dimensional uniform distribution:

$$p_0(\mu) = \mathcal{U}(\log(\mu_{0,min}), \log(\mu_{0,max})), \quad (10a)$$

with $\mu_{0,min}$ and $\mu_{0,max}$ of the three parameters in the three luminance conditions specified in Table 1.

The weakly informative prior distribution of Σ , $p_0(\Sigma)$, was specified by a 9×9 precision matrix Ω with a Wishart distribution:

$$p_0(\Omega) = \mathcal{W}(\Sigma_{BIP}^{-1}/\nu, \nu), \quad (10b)$$

$$p_0(\Sigma) = p_0(\Omega^{-1}), \quad (10c)$$

where the degrees of freedom $\nu = 9$, and the expected mean, Σ_{BIP}^{-1} , was based on the covariance matrix of the estimated CSF parameters Σ_{BIP} across all the subjects and luminance conditions from the BIP procedure.

The weakly informative prior distribution of ϕ_j , $p_0(\phi_j)$, was specified with a 3×3 precision matrix Λ_j with a Wishart distribution:

$$p_0(\Lambda_j) = \mathcal{W}(\phi_{BIP,j}^{-1}/\nu_j, \nu_j), \quad (10d)$$

$$p_0(\phi_j) = p_0(\Lambda_j^{-1}), \quad (10e)$$

where the degrees of freedom $\nu_j = 3$, and the expected mean, $\phi_{BIP,j}^{-1}$, was based on the average covariance matrix $\phi_{BIP,j}$ computed from the estimated CSF

parameters across all the subjects in luminance condition j from the BIP procedure.

The R (R Core Team, 2020) function `autorun.jags` in JAGS (Plummer, 2003) was used to compute representative samples of the posterior distributions of θ_{ij1} (3 parameters/condition \times 3 conditions \times 112 subjects = 1008 parameters), $\rho_{i,1:J}$ (9 parameters \times 112 subjects = 1008 parameters), ϕ_j (6 parameters/condition \times 3 conditions = 18 parameters), μ (9 parameters), and Σ (45 parameters) in three Markov Chain Monte Carlo (MCMC) chains. The MCMC is an algorithm used to efficiently sample the joint posterior distribution (Kruschke, 2015). It started at a randomly selected position in the 2088-dimensional parameter space. In each step, one of the 2088 parameters was selected randomly. The one-dimensional conditional posterior probability distribution of the selected parameter was evaluated by fixing the values of all the other 2087 parameters at the current position. A new value of the selected parameter was chosen based on the one-dimensional conditional probability distribution (Equation 8). By reiterating this process, the probability of visiting a location in the random walk approximated the joint posterior distribution of all the 2088 parameters in Equation 8. These steps were re-iterated until the convergence criterion was reached.

Gelman and Rubin's diagnostic (Gelman & Rubin, 1992), the ratio of between-chain and within-chain variances, was used to quantify the convergence between different MCMC chains. The convergence criterion was set at 1.05 for all parameter estimates. After the convergence criterion was met, the program terminated when 1,000,000 total samples were generated in each MCMC chain. There were 10,000 of the 1,000,000 samples that were stored (thinning ratio = 100) to ensure at least 10,000 effective samples of X in subsequent analysis.

Statistical analysis

Goodness of fit

Bayesian predictive information criterion (BPIC; Ando 2007; Ando 2011) was used to quantify the goodness of fit to the trial-by-trial data. The BPIC quantifies the likelihood of the data based on the joint posterior distribution of the parameters of the model and penalizes model complexity.

Posterior distributions of the area under log CSF

The posterior distributions of AULCSF were constructed by computing the AULCSFs from samples of the corresponding posterior distributions of θ_{ij1} from the HBM and BIP.

***d'*: Between-condition discriminability of distributions**

Discriminability d' quantifies the signal (mean separation) to noise (variability) ratio of two probability distributions. We used the difference distribution between conditions to compute d' . Each sample in the difference distribution represented the difference between two randomly drawn samples from the corresponding distributions.

For a one-dimension difference distribution, d' is defined as (Green & Swets, 1966):

$$d' = \sqrt{2} \frac{\Delta}{\sigma}, \quad (11)$$

where Δ is the mean separation, and σ is the standard deviation of the difference distribution.

For a multidimensional difference distribution, d' is defined as (Ashby & Townsend, 1986):

$$d' = \sqrt{2\Delta * cov(\Delta)^{-1} * \Delta^T}, \quad (12)$$

where Δ and $cov(\Delta)$ are the mean separation and covariance matrix of the difference distribution, $cov(\Delta)^{-1}$ is the inverse of $cov(\Delta)$, Δ^T is the transpose of Δ , and $*$ represents matrix multiplication.

Statistical tests on CSF parameters and AULCSF

We compared the mean (expected value) and variance of the posterior distributions of θ_{ij1} from the HBM and BIP using Hotelling's T-squared test (Anderson, 2003) in R (R Core Team, 2020; Nordhausen, Sirkia, Oja, & Tyler, 2018). We also compared the correlation coefficients of pairs of CSF parameters from the two methods with paired t -test.

To quantify the between-condition discriminability across subjects, we compared the means of the posterior distributions of θ_{ij1} and AULCSF between pairs of experimental conditions from each method with Hotelling's T-squared test and paired t -test, respectively.

Simulation

To compare the accuracy and precision of the BIP and HBM estimates, we conducted a simulation study to investigate the bias and variability of the estimated CSF parameters for each test. The dataset consisted of 336 qCSF tests (112 subjects \times 3 conditions). $\theta_{1:J,1:J,1}$ of the simulated tests were a random sample from the posterior distribution of $\tau_{1:J,1:J}$ obtained from the HBM fit to the real data. Each qCSF test consisted of 150 trials, identical to the real experiment (Hou et al., 2016), with the trial-by-trial responses determined by the CSF parameters of the simulated subject

(Equations 1 to 3). Both the HBM and BIP were fit to the simulated dataset. The mean of the posterior distribution of θ_{ij1} was used as the best estimate for each test. The bias, root mean square error (RMSE), variance, d' , and t statistics were computed based on the posterior distributions of θ_{ij1} from both methods.

Results

Goodness of fit

The BPIC for the BIP and HBM were 34886 and 34225, respectively, indicating that the HBM fit the data better than the BIP. Figure 3 shows the estimated CSFs of one subject from the BIP and HBM in three luminance conditions.

Many “image-computable” models have used the CSF as the front-end filter on actual images to predict human performance in image processing and object recognition (Chung, Legge, & Tjan, 2002; Malo, Pons, Felipe, & Artigas, 1997; Schütt, & Wichmann, 2017; Watson, 2000; Watson & Ahumada Jr, 2005; Watson, & Malo, 2002). Although recent studies have suggested that a more comprehensive model may require additional parameters related to non-linearities in the visual system (Chen, Hou, Yan, Zhang, Xi, Zhou, Lu, & Huang, 2014; Hou, Lu, & Huang, 2014), the CSF filtered images nevertheless provide an excellent demonstration of human visual processing. To illustrate the differences between the CSF estimates from the BIP and HBM and their implications for image-computable models (Figure 4), we applied the mean - SD, mean, and mean + SD CSFs from the BIP and HBM in the high luminance condition to filter a letter K (Lu & Doshier, 2013). Although the mean CSFs from the two methods are very similar and generated very similar filtered K's, the CSFs from the BIP exhibited much larger uncertainty.

Posterior distributions from the HBM

Figure 5 shows the three-dimensional posterior distributions of hyperparameters η (marginalized), $\tau_{i,1:J}$ for one individual, and $\theta_{i,1:J,1}$ for one individual in one test from the HBM.

Population level

Table 2 shows the mean and covariance matrix of η . The correlation coefficients were positive and significant between all three pairs of experimental conditions. Table 3 shows the d' s of η between the three pairs of experimental conditions. In the HBM, the posterior distributions of η constrained $\tau_{i,1:J}$. The large d' s of the posterior distributions of η between

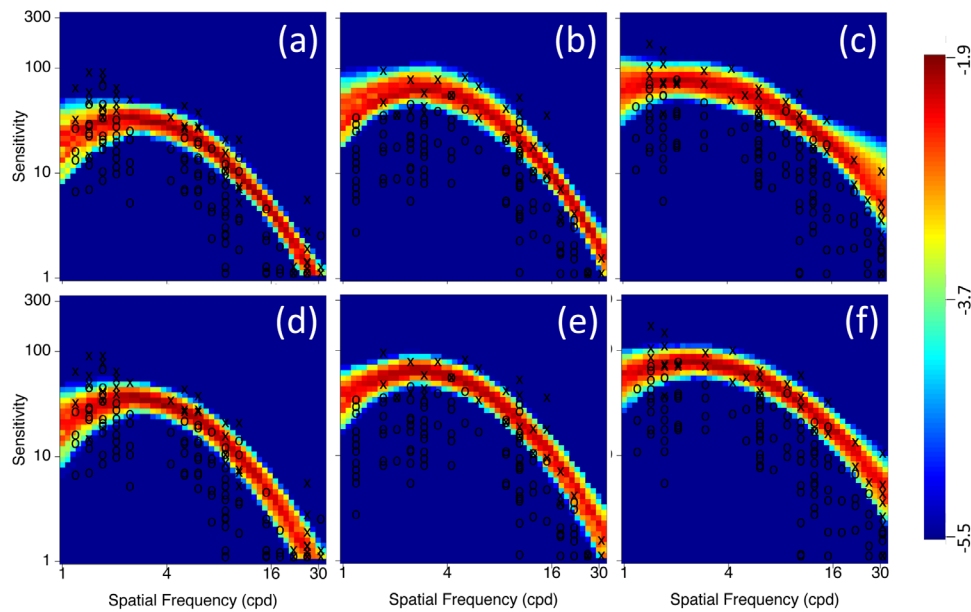


Figure 3. Estimated CSFs of one subject from the BIP (a, b, c) and HBM (d, e, f) methods in three luminance conditions (a, d) $L = 2.62 \text{ cd/m}^2$, (b, e) $M = 20.4 \text{ cd/m}^2$, and (c, f) $H = 95.4 \text{ cd/m}^2$. Color map indicates \log_{10} probability density.

different experimental conditions indicated that the posterior distributions of η provided strong constraints on $\tau_{i,1:J}$.

Individual level

Table 4 shows the average covariance matrix of $\tau_{1:I,1:J}$ across all 112 individuals and experimental conditions. Figure 5(b) illustrates the three-dimensional posterior distributions of $\tau_{i,1:J}$ for one individual in all three luminance conditions. Table 3 shows the average d 's of τ_{ij} between the three pairs of experimental conditions. In the HBM, the posterior distributions of $\tau_{i,1:J}$ constrained $\theta_{i,1:J,1}$. The large d 's of the posterior distributions of $\tau_{i,1:J}$ between different experimental

conditions indicated that the posterior distributions of $\tau_{i,1:J}$ provided strong constraints on $\theta_{i,1:J,1}$.

Test level

We computed the mean, covariance, and correlation coefficient based on the estimated test-level CSF parameters $\theta_{1:I,1:J,1}$ in the three luminance conditions from the HBM and compared them with the results from the BIP.

The means of the posterior distributions of θ_{ij1} from the HBM and BIP were significantly different ($t^2(9,103) = 5.34, p < 0.001$), and the average variance of the estimated CSF parameters from the HBM (mean = $0.00139 \log_{10}$ units; range = 0.00030 to $0.00739 \log_{10}$ units) was 65.8% less than that from the BIP (mean = $0.00407 \log_{10}$ units; range = 0.00035 to $0.11893 \log_{10}$ units) ($t^2(9,103) = 109, p < 0.001$), consistent with the well-known variance shrinkage effect of the HBM (Kruschke, 2015).

Figures 6 and 7 show histograms of the difference between the expected values of θ_{ij1} , and the standard deviation ($\text{SD} = \sqrt{\text{variance}}$) of θ_{ij1} from the BIP and HBM. Whereas most of the differences between the expected values of θ_{ij1} from the two methods were small (mean absolute difference = $0.027 \log_{10}$ units), there were thirteen instances (out of a total of $3 \text{ parameters} \times 3 \text{ conditions} \times 112 \text{ subjects} = 1008$) in which the absolute difference was greater than $0.2 \log_{10}$ units (range = 0.200 to $0.676 \log_{10}$ units). The discrepancies were associated with large variances of the BIP estimates in those instances: their average variance of $0.065 \log_{10}$ units (range = 0.027 to $0.119 \log_{10}$ units)

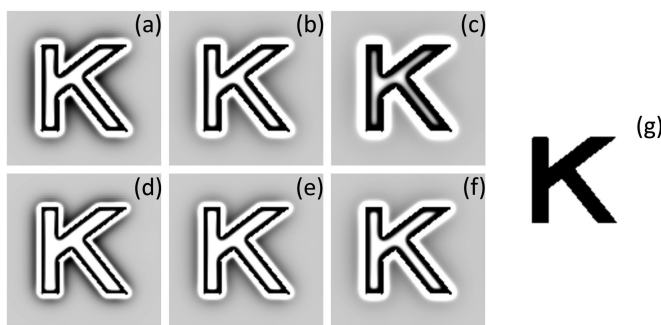


Figure 4. Visualization of the CSF estimates from the BIP and HBM. Filtered letter K by the mean - SD (a, d), mean (b, e) and mean + SD (c, f) CSFs from the BIP (a, b, c) and HBM (d, e, f) in the high luminance condition. The original image is shown in (g).

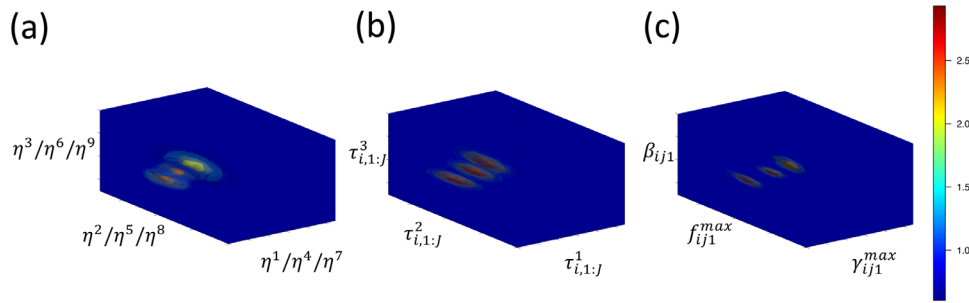


Figure 5. Three-dimensional posterior distributions of η (marginalized) (a), $\tau_{i,1:j}$ for one individual (b), and $\theta_{i,1:j,1}$ (c) for one individual in one test in the HBM. The colors represent log10 probability density.

		Covariance									
		L			M			H			
Parameter		η^1	η^2	η^3	η^4	η^5	η^6	η^7	η^8	η^9	Mean
L	η^1	6.2E-03	1.1E-03	4.4E-05	4.7E-03	1.9E-03	-4.7E-04	4.2E-03	-2.3E-04	8.4E-04	1.564
	η^2	1.1E-03	9.5E-03	-1.9E-03	3.2E-03	4.5E-03	-5.3E-04	2.6E-03	1.8E-03	1.0E-03	0.319
	η^3	4.4E-05	-1.9E-03	1.1E-03	3.0E-05	-6.9E-04	5.6E-04	-1.2E-04	-5.4E-04	5.3E-04	0.462
M	η^4	4.7E-03	3.2E-03	3.0E-05	6.9E-03	1.7E-03	-1.1E-04	5.4E-03	-1.5E-04	1.2E-03	1.729
	η^5	1.9E-03	4.5E-03	-6.9E-04	1.7E-03	6.3E-03	-1.7E-03	2.6E-03	1.9E-03	2.9E-04	0.387
	η^6	-4.7E-04	-5.3E-04	5.6E-04	-1.1E-04	-1.7E-03	1.2E-03	-6.8E-04	-8.1E-04	9.2E-04	0.478
H	η^7	4.2E-03	2.6E-03	-1.2E-04	5.4E-03	2.6E-03	-6.8E-04	8.5E-03	-2.7E-03	1.3E-03	1.810
	η^8	-2.3E-04	1.8E-03	-5.4E-04	-1.5E-04	1.9E-03	-8.1E-04	-2.7E-03	1.0E-02	-3.7E-03	0.385
	η^9	8.4E-04	1.0E-03	5.3E-04	1.2E-03	2.9E-04	9.2E-04	1.3E-03	-3.7E-03	2.9E-03	0.520

Table 2. Mean and covariance of η . H, high; L, low; M, medium.

was 16 times the mean variance (0.00407 log10 units) of θ_{ij1} in the BIP procedure, suggesting that BIP did not converge well in those cases. On the other hand, the HBM generated more precise estimates with on average a 93.7% reduction of variance (mean = 0.00407 log10 units; range = 0.00139 to approximately 0.00681 log10 units) compared to the BIP in the 13 cases by incorporating data from all the subjects and conditions in a single model.

Table 5 lists the average correlation coefficients between θ_{ij1} in pairs of luminance conditions across subjects from the HBM and BIP procedures. All correlations were negative, with the strongest between f^{max} and β . Across all the subjects, 97.4% and 97.9% of the correlation coefficients from the BIP and HBM were statistically significant, respectively. Although the paired t -test showed that the correlation coefficients between

Hyperparameter	M-L	H-L	H-M
η	8.1	10.2	5.6
τ_{ij}	11.6	11.6	5.5

Table 3. d' of η and average d' of τ_{ij} between pairs of luminance conditions. H, high; L, low; M, medium.

γ^{max} and β in the high luminance condition ($p = 0.003$) and between f^{max} and β in all three luminance conditions ($p < 0.001$) from the two procedures were significantly different, the magnitudes of the differences were very small and probably not of practical importance.

Table 6 shows the average d' s of θ_{ij1} and AULCSF between pairs of luminance conditions across all the subjects. Averaged across the three pairs, the AULCSF d' from the HBM was 33.5% greater than that from the BIP. Compared to AULCSF, incorporating information from the three-dimensional joint distributions of θ_{ij1} led to an average d' increase of 66.6% for the BIP and 51.7% for the HBM. Compared to the AULCSF d' from the BIP, using θ_{ij1} in the HBM increased d' by 103.3% across the three pairs of luminance conditions.

Statistics on θ_{ij1} and AULCSF across individuals

Table 7 shows $\sqrt{t^2(3, 109)}$ and $t(111)$ of the means of θ_{ij1} and AULCSF among the three pairs of experimental conditions. The HBM generated larger t values than the BIP for both θ_{ij1} and AULCSF in all pairs of experimental conditions. Averaged across the

Parameter	L			M			H		
	$\tau_{1:,1}^1$	$\tau_{1:,1}^2$	$\tau_{1:,1}^3$	$\tau_{1:,2}^1$	$\tau_{1:,2}^2$	$\tau_{1:,2}^3$	$\tau_{1:,3}^1$	$\tau_{1:,3}^2$	$\tau_{1:,3}^3$
$\tau_{1:,j}^1$	2.6E-03	-9.8E-04	-1.0E-06	2.2E-03	-8.0E-04	-3.2E-05	3.4E-03	-1.9E-03	5.5E-05
$\tau_{1:,j}^2$	-9.8E-04	4.6E-03	-1.8E-03	-8.0E-04	3.6E-03	-1.4E-03	-1.9E-03	8.0E-03	-2.9E-03
$\tau_{1:,j}^3$	-1.0E-06	-1.8E-03	8.7E-04	-3.2E-05	-1.4E-03	7.2E-04	5.5E-05	-2.9E-03	1.4E-03

Table 4. Average covariance matrix of $\tau_{1:,1:j}$. H, high; L, low; M, medium.

three pairs, $t(111)$ of AULCSF and $\sqrt{t^2(3, 109)}$ of θ_{ij1} from the HBM were 51.2% and 49.6% greater than those from the BIP.

Simulation

The HBM accurately and precisely recovered θ_{ij1} in the simulation, with very small bias ($\gamma^{max}, f^{max}, \beta$: 0.0028, -0.0091, and 0.0023 log10 units), RMSE (0.0373 log10 units), and average variance (0.00149 log10 units). In comparison, the BIP exhibited lower accuracy and precision (bias = $\gamma^{max}, f^{max}, \beta$: 0.0147, -0.0395, and 0.0118 log10 units; RMSE = 0.0673 log10 units; average variance = 0.00428 log10 units).

Discussion

The HBM provides a general framework that can be adapted to different experiment designs. In this paper, we developed a three-level HBM to account for CSF data of 112 subjects in a single-factor (luminance), multi-condition (3 luminance conditions), and within-subject experimental design. We applied the HBM to quantify the joint distribution of CSF parameters and hyperparameters at the population, individual, and test levels and compared the performance of the model with that of the BIP. The HBM generated more precise estimates for each test than the BIP by incorporating information across subjects and conditions to constrain the estimates. The increased precision led to increased

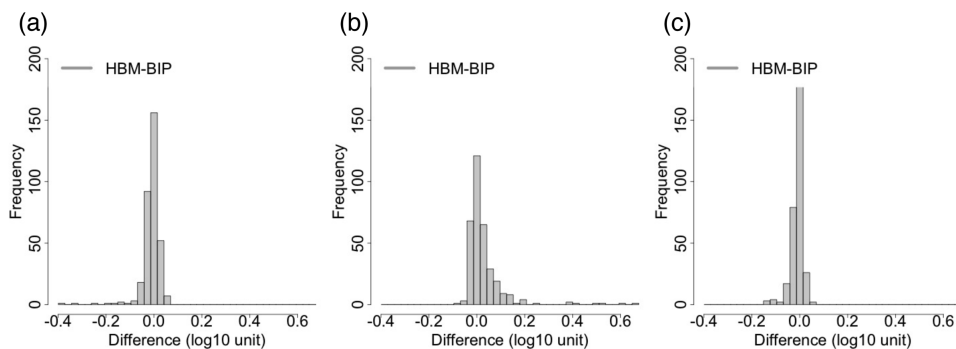


Figure 6. Histograms of the difference between the expected values of θ_{ij1} from the HBM and BIP. (a) γ_{ij1}^{max} ; (b) f_{ij1}^{max} ; and (c) β_{ij1} .

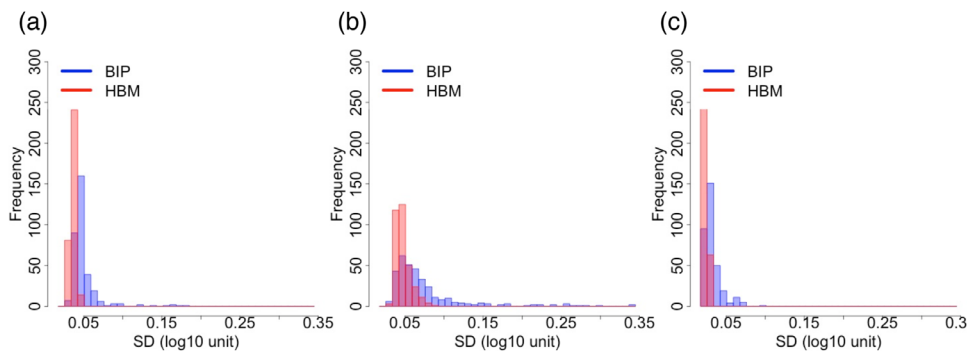


Figure 7. Histograms of the standard deviation (SD) of θ_{ij1} from the HBM and BIP procedures. (a) γ_{ij1}^{max} ; (b) f_{ij1}^{max} ; and (c) β_{ij1} .

Method	Parameter pair		Condition		
			L	M	H
BIP	γ^{max}	f^{max}	-0.185	-0.134	-0.214
	γ^{max}	β	-0.072	-0.145	-0.078
	f^{max}	β	-0.920	-0.910	-0.904
HBM	γ^{max}	f^{max}	-0.197	-0.169	-0.174
	γ^{max}	β	-0.095	-0.143	-0.150
	f^{max}	β	-0.899	-0.891	-0.887

Table 5. Average correlations between θ_{ij1} in pairs of luminance conditions. H, high; L, low; M, medium.

Method	Metric	M-L	H-L	H-M
BIP	AULCSF	7.5	10.8	3.4
	θ_{ij1}	10.8	16.0	7.0
HBM	AULCSF	10.1	13.5	4.8
	θ_{ij1}	13.4	19.2	8.6

Table 6. Average d' of θ_{ij1} and AULCSF between pairs of luminance conditions. H, high; L, low; M, medium.

d' s of AULCSF and CSF parameters between different experimental conditions at the test level for each subject, and bigger statistical differences across subjects. Relative to the BIP, the HBM increased the average d' s of AULCSF and θ_{ij1} between conditions at the test level by 24.5% and 20.5%, and the corresponding $t(111)$ and $\sqrt{t^2(3, 109)}$ by 51.2% and 49.6%, respectively. Simulations also showed that the HBM generated accurate and precise CSF parameter estimates.

The HBM generated larger d' and t statistics at the test level because it reduced the variance of θ_{ij1} by 65.8% relative to the BIP (0.00139 vs. 0.00407 log10 units). In addition, the 13 instances in which the absolute difference of θ_{ij1} from the HBM and BIP were greater than 0.2 log10 units further demonstrated the benefit of incorporating information across tests, subjects, and conditions in the HBM (Kruschke, 2015; Rouder & Lu, 2005; Rouder, Sun, Speckman, Lu, & Zhou, 2003). In those cases, the variances of the BIP estimates were very large (16 times the mean variance), suggesting that

Method	Metric	M-L	H-L	H-M
BIP	AULCSF	41.78	45.35	21.01
	θ_{ij1}	27.05	34.90	17.87
HBM	AULCSF	63.88	62.78	34.11
	θ_{ij1}	40.99	48.35	28.38

Table 7. $\sqrt{t^2(3, 109)}$ and $t(111)$ between means of θ_{ij1} and AULCSF. H, high; L, low; M, medium.

the BIP did not converge well. On the other hand, the HBM generated much more precise CSF estimates for each test by incorporating data across subjects and conditions in a single model. The ability of the HBM to generate more precise estimates from insufficient or poor-quality data can be quite valuable in clinical trials.

The HBM can be used to conduct two types of power analyses. First, a replication power analysis computes the power of different sample sizes in replicated experiments with the exact same experimental design (Kruschke, 2015). Simulated data for new subjects can be generated from the posterior distributions of the hyperparameters based on the HBM fit to the existing dataset, just as we did in the simulation. In this case, the original data of 112 subjects shall be combined with the simulated data to compute the power for each new sample size. A more interesting application of the HBM is in prospective power analysis (Kruschke, 2015). In that case, no data have been collected for a new experiment; simulated data of the new experiment must be based on the generative model constructed based on results from a different experiment. An HBM for the new experimental design is then constructed and fit to the simulated data. Therefore, data from existing studies can only be used as prior knowledge and cannot be combined with the simulated data.

A certain sample size is required for the joint posterior distribution of the CSF hyperparameters and parameters in the HBM to become stable. This can be done by evaluating the stability of the estimates with different sample sizes. Relative to the noninformative diffuse prior used in the BIP, Gu, Kim, Hou, Lesmes, Pitt, Lu, and Myung (2016) showed that (1) an informative prior from the HBM fit to as few as five subjects could provide significant improvement in qCSF measurements of new subjects in the hierarchical adaptive design optimization (HADO) procedure, (2) prior constructed from larger samples further improved the accuracy and precision of the estimation, and (3) the improvement stabilized when the sample size was about 30. Simulation studies are necessary to determine the minimum sample size required for the HBM to converge for a given experiment.

Although the MCMC algorithm automatically selected by the JAGS (Plummer, 2003) provided an efficient sampling method, the weakly informative priors of the covariance matrices at both the subject and population levels were very helpful. With these priors, it took about 54 hours for each MCMC chain to generate at least 10,000 effective samples (Kruschke, 2015) for all parameters on a computer with eight cores. Fifty-four hours are practical given that the actual physical time decreases with increasing number of CPUs used in parallel computation (Kruschke, 2015). On the other hand, the HBM using diagonal covariance

matrices as the priors took 18% longer (63.5 hours) to generate 10,000 effective samples for all parameters, and did not converge, as indicated by the larger variances of the estimated CSF parameters at both the subject (0.00298 log₁₀ units and 114% increase) and population levels (0.212 log₁₀ units and 2873% increase), although the Gelman and Rubin's diagnostic was below 1.05 for all parameters, which is based on the ratio of between-chain and within-chain variances but not the magnitudes of the variances. The effects of priors on covariance estimation were consistent with previous studies (Hobert & Casella, 1996; Rouder et al., 2003).

Although the HBM in the current study was developed to account for group differences in a within-subject design, HBM-based approaches can be developed to detect deviations in individual patients belonging to different subpopulations (e.g. healthy versus different stages of an eye disease, or different eye diseases) using different tests (null hypothesis significance testing versus estimation of magnitude/effect size) in both frequentist and Bayesian approaches (Kruschke & Liddell, 2018). The HADO method (Gu, Kim, Hou, Lesmes, Pitt, Lu, & Myung, 2016; Kim, Pitt, Lu, Steyvers, & Myung, 2014) provides a potential framework. HADO uses an informative prior obtained from all previously tested subjects with the same CSF characteristics (e.g. testing in the same luminance condition). It took 20 to 30 trials for the qCSF method with an uninformative diffuse prior to achieve the same initial precision level of the HADO procedure with the informative prior. Moreover, HADO with a mixture prior that represent a wide range of CSF properties (e.g. across different luminance conditions) still achieved higher precision than qCSF with an uninformative prior, and could mitigate the problem of mis-specified prior and improve the qCSF method on testing individual subjects. The joint posterior distributions of the hyperparameters at the population and individual levels from HBM can provide informative priors within the HADO framework for new individuals and repeated tests of the same individual, respectively. Furthermore, the HBM can be extended to model additional covariance between parameters of different measurements (e.g. CSF and visual acuity) in a joint modeling approach (Palestro, Bahg, Sederberg, Lu, Steyvers, & Turner, 2018; Turner, Forstmann, Wagenmakers, Brown, Sederberg, & Steyvers, 2013) to account for multiple test results of multiple subjects and conditions, and potentially further increase statistical power in detecting changes of functional vision in normal and clinical populations. Therefore, the HBM framework can be used to take advantage of all available information at different levels to enable sensitive detection of CSF changes, and thereby improve patient care and clinical trials with increased statistical power.

Conclusions

In this paper, we developed a three-level HBM to account for CSF data of 112 subjects in a within-subject, single-factor (luminance), multi-condition (3 luminance conditions) experimental design. The HBM was used to compute the joint distribution of CSF parameters and hyperparameters at population, individual, and test levels to fully utilize information across levels to accurately estimate the CSF in each test. Relative to the BIP, the HBM increased the average d 's of AULCSF and θ_{ij1} between conditions at the test level by 24.5% and 20.5 %, and the corresponding t statistics by 51.2% and 49.6%, respectively. Future research will further evaluate the potential value of the HBM for analyzing clinical changes in contrast sensitivity, whether in individual patients or groups in clinical trials.

Keywords: contrast sensitivity function, quick contrast sensitivity function (qCSF), hierarchical Bayesian model (HBM), d 's

Acknowledgments

Supported by the National Eye Institute (EY021553 and EY017491 to Z.L.), and by the Qianjiang Talent Project of Zhejiang Province (QJD1803028 to H.F.).

Commercial relationships: L.A.L. and Z.L.L. have intellectual property interests in methods for measuring and applying contrast sensitivity functions (US 7938538, WO2013170091, and PCT/US2015/028657), and equity interest in Adaptive Sensory Technology, Inc. (San Diego, CA). In addition, L.A.L. has an intellectual property interest in methods for quantitative visual acuity testing (US 10758120B2) and holds employment in AST.

Corresponding author: Zhong-Lin Lu.

Email: zhonglin@nyu.edu.

Address: Center for Neural Science and Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA.

Footnote

¹Hou et al., (2016) used a four-parameter CSF model with a low frequency truncation level δ as the fourth parameter (Lesmes et al., 2010; Watson & Ahumada, 2005). Because of the range of low-frequencies that were tested in this dataset, we used the three-parameter CSF model without δ because it was not very well constrained. The three-parameter log-parabola CSF also reduced the total number of parameters and the complexity of the HBM.

References

- Ahn, W.-Y., Krawitz, A., Kim, W., Busmeyer, J. R., & Brown, J. W. (2011). A Model-Based fMRI Analysis with Hierarchical Bayesian Parameter Estimation. *Journal of Neuroscience, Psychology, and Economics*, 4(2), 95–110.
- Anderson, T. W. (2003). *An introduction to multivariate analysis*. Hoboken, N.J.: Wiley-Interscience.
- Ando, T. (2007) Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, 94, 443–458.
- Ando, T. (2011) Predictive Bayesian Model Selection. *American Journal of Mathematical and Management Sciences*, 31, 13–38.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of Perceptual Independence. *Psychological Review*, 93(2), 154–179.
- Bellmann, C., Unnebrink, K., Rubin, G. S., Miller, D., & Holz, F. G. (2003). Visual acuity and contrast sensitivity in patients with neovascular age-related macular degeneration. Results from the Radiation Therapy for Age-Related Macular Degeneration (RAD-) Study. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 241, 968–974.
- Bellucci, R., Scialdone, A., Buratto, L., Morselli, S., Chierigo, C., & Criscuoli, A. et al. (2005). Visual acuity and contrast sensitivity comparison between Tecnis and AcrySof SA60AT intraocular lenses: A multicenter randomized study. *Journal of Cataract & Refractive Surgery*, 31, 712–717.
- Borm, G. F., Fransen, J., & Lemmens, W. A. J. G. (2007). A Simple Sample Size Formula for Analysis of Covariance in Randomized Clinical Trials. *Journal of Clinical Epidemiology*, 60(12), 1234–1238.
- Bradley, A., Hook, J., & Haeseker, J. (1991). A comparison of clinical acuity and contrast sensitivity charts: Effect of uncorrected myopia. *Ophthalmic and Physiological Optics*, 11, 218–226.
- Buhren, J., Terzi, E., Bach, M., Wesemann, W., & Kohlen, T. (2006). Measuring contrast sensitivity under different lighting conditions: Comparison of three tests. *Optometry & Vision Science*, 83, 290–298.
- Chen, G., Hou, F., Yan, F.-F., Zhang, P., Xi, J., & Zhou, Y. et al. (2014). Noise Provides New Insights on Contrast Sensitivity Function. *PLoS One*, 9(3), e90579.
- Chung, S. T. L., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research*, 42, 2137–2152.
- Clayton, D., & Hills, M. (1993) *Statistical models in epidemiology*. Oxford, UK: Oxford University Press.
- Daniels, M. J., & Kass, R. E. (1999). Nonconjugate Bayesian Estimation of Covariance Matrices and Its Use in Hierarchical Models. *Journal of the American Statistical Association*, 94(448), 1254–1263.
- Edwards, W., Lindman, H., & Savage, L.J. (1963) Bayesian statistical inference for psychological research. *Psychological Review*, 70(3), 193–242.
- Egbewale, B. E., Lewis, M., & Sim, J. (2014). Bias, Precision and Statistical Power of Analysis of Covariance in the Analysis of Randomized Trials with Baseline Imbalance: A Simulation Study. *BMC Medical Research Methodology*, 14, 49.
- Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, 21, 1041–1053.
- Gelman, A., & Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457–511.
- Ginsburg, A. P. (1981). Spatial filtering and vision: Implications for normal and abnormal vision. In L. Proenz, J. Enoch, & A. Jampolsky (Eds.), *Clinical applications of visual psychophysics* (pp. 70–106). Cambridge, UK: Cambridge University Press.
- Ginsburg, A. P. (2003). Contrast Sensitivity and Functional Vision. *International Ophthalmology Clinics*, 43(2), 5–15.
- Ginsburg, A. P. (2006). Contrast sensitivity: Determining the visual quality and function of cataract, intraocular lenses and refractive surgery. *Current Opinion in Ophthalmology*, 17, 19–26.
- Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: John Wiley & Sons.
- Gu, H., Kim, W., Hou, F., Lesmes, L. A., Pitt, M. A., Lu, Z.-L., . . . Myung, J. I. (2016). A Hierarchical Bayesian Approach to Adaptive Vision Testing: A Case Study with the Contrast Sensitivity Function. *Journal of Vision*, 16(6), 15.
- Haymes, S. A., Roberts, K. F., Cruess, A. F., Nicolela, M. T., LeBlanc, R. P., & Ramsey, M. S. et al. (2006). The letter contrast sensitivity test: Clinical evaluation of a new design. *Investigative Ophthalmology & Visual Science*, 47, 2739–2745.
- Hess, R. F. (1981). Application of contrast-sensitivity techniques to the study of functional amblyopia. In L. Proenz, J. Enoch, & A. Jampolsky (Eds.), *Clinical applications of visual psychophysics* (pp. 11–41). Cambridge, UK: Cambridge University Press.

- Hobert, J. P., & Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, *91*, 1461–1473
- Hohberger, B., Laemmer, R., Adler, W., Juenemann, A. G., & Horn, F. K. (2007). Measuring contrast sensitivity in normal subjects with OPTEC 6500: Influence of age and glare. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*, 1805–1814.
- Hou, F., Huang, C.-B., Lesmes, L. A., Feng, L.-X., Tao, L., Zhou, Y.-F., . . . Lu, Z.-L. (2010). qCSF in clinical application: Efficient characterization and classification of contrast sensitivity functions in amblyopia. *Investigative Ophthalmology & Visual Science*, *51*, 5365–5377.
- Hou, F., Lesmes, L. A., Bex, P., Dorr, M., & Lu, Z.-L. (2015). Using 10AFC to further improve the efficiency of the qCSF method. *Journal of Vision*, *15*(9):2, 1–18.
- Hou, F., Lesmes, L. A., Kim, W., Gu, H., Pitt, M. A., Myung, J. I., . . . Lu, Z.-L. (2016). Evaluating the Performance of the QCSF Method in Detecting Contrast Sensitivity Function Changes. *Journal of Vision*, *16*(6), 18.
- Hou, F., Lu, Z.-L., & Huang, C.-B. (2014). The external noise normalized gain profile of spatial vision. *Journal of Vision*, *14*(13):9, 1–14.
- Jia, W., Zhou, J., Lu, Z.-L., Lesmes, L. A., & Huang, C.-B. (2015). Discriminating Anisometropic Amblyopia from Myopia Based on Interocular Inhibition. *Vision Research*, *114*, 135–141.
- Jindra, L. F., & Zemon, V. (1989). Contrast Sensitivity Testing - a More Complete Assessment of Vision. *Journal of Cataract and Refractive Surgery*, *15*(2), 141–148.
- Joltikov, K. A., de Castro, V. M., Davila, J. R., Anand, R., Khan, S. M., Farbman, N., . . . Jackson, G. R. et al. (2017). Multidimensional functional and structural evaluation reveals neuroretinal impairment in early diabetic retinopathy. *Investigative Ophthalmology & Visual Science*, *58*, BIO277–BIO290.
- Kalia, A., Lesmes, L. A., Dorr, M., Gandhi, T., Chatterjee, G., & Ganesh, S. et al. (2014). Development of pattern vision following early and extended blindness. *Proceedings of the National Academy of Sciences, USA*, *111*, 2035–2039.
- Kelly, D. H., & Savoie, R. E. (1973). A study of sinewave contrast sensitivity by two psychophysical methods. *Perception & Psychophysics*, *14*, 313–318.
- Kim, W., Pitt, M. A., Lu, Z.-L., Steyvers, M., & Myung, J. I. (2014). A Hierarchical Adaptive Approach to Optimal Experimental Design. *Neural Computation*, *26*(11), 2465–2492.
- Kleiner, R. C., Enger, C., Alexander, M. F., & Fine, S. L. (1988). Contrast sensitivity in age-related macular degeneration. *Archives of Ophthalmology*, *106*, 55–57.
- Klotzke, K., & Fox, J.-P. (2019). Bayesian Covariance Structure Modeling of Responses and Process Data. *Frontiers in Psychology*, *10*, 1675.
- Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian Adaptive Estimation of Psychometric Slope and Threshold. *Vision Research* *39*(16), 2729–2737.
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan*. San Diego, CA: Academic Press.
- Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206.
- Kuss, M., Jäkel, F., & Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *Journal of Vision*, *5*(5), 8.
- Lee, M. D. (2006). A Hierarchical Bayesian Model of Human Decision-Making on an Optimal Stopping Problem. *Cognitive Science*, *30*(3), 1–26.
- Lee, M. D. (2011). How Cognitive Modeling Can Benefit from Hierarchical Bayesian Models. *Journal of Mathematical Psychology*, *55*(1), 1–7.
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian Inference in the Visual Cortex. *Journal of the Optical Society of America A - Optics Image Science and Vision*, *20*(7), 1434–1448.
- Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *4*, 391–404.
- Lesmes, L. A., Jackson, M.L., & Bex, P. (2013). Visual function endpoints to enable dry AMD clinical trials. *Drug Discovery Today: Therapeutic Strategies*, *10*(1), e43–e50.
- Lesmes, L. A., Lu, Z.-L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The qCSF method. *Journal of Vision*, *10*(3):17, 1–21.
- Lesmes, L. A., Wallis, J., Jackson, M. L., & Bex, P. (2013). The reliability of the qCSF method for contrast sensitivity assessment in low vision. *Investigative Ophthalmology & Visual Science*, *54*, 2762. [Abstract]
- Lesmes, L. A., Wallis, J., Lu, Z.-L., Jackson, M. L., & Bex, P. J. (2012). Clinical application of a novel contrast sensitivity test to a low vision population:

- The qCSF method. *ARVO Meeting Abstracts*, 53, 4358.
- Levi, D. M., & Li, R. W. (2009). Improving the performance of the amblyopic visual system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 399–407.
- Li, F.-F., & Perona, P. (2005). A Bayesian Hierarchical Model for Learning Natural Scene Categories. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 2*, pp. 524–531.
- Lin, S., Mihailovic, A., West, S. K., Johnson, C. A., Friedman, D. S., Kong, X., . . . Ramulu, P. Y. (2018). Predicting Visual Disability in Glaucoma With Combinations of Vision Measures. *Translational Vision Science & Technology*, 7(2), 22.
- Loshin, D. S., & White, J. (1984). Contrast sensitivity. The visual rehabilitation of the patient with macular degeneration. *Archives of Ophthalmology*, 102, 1303–1306.
- Lu, Z. L., & Doshier, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 16, 764–778.
- Lu, Z.-L., & Doshier, B.A. (2013). *Visual Psychophysics: From Laboratory to Theory*. Cambridge, MA: MIT Press.
- Malo, J., Pons, A.M., Felipe, A., & Artigas, J.M. (1997). Characterization of the human visual system threshold performance by a weighting function in the Gabor domain. *Journal of Modern Optics*, 44(1), 127–148.
- Marmor, M. F. (1986). Contrast sensitivity versus visual acuity in retinal disease. *British Journal of Ophthalmology*, 70, 553–559.
- Merkle, E. C., Smithson, M., & Verkuilen, J. (2011). Hierarchical Models of Simple Mechanisms Underlying Confidence in Decision Making. *Journal of Mathematical Psychology*, 55(1), 57–67.
- Midena, E., Degli Angeli, C., Blarzino, M. C., Valenti, M., & Segato, T. (1997). Macular function impairment in eyes with early age-related macular degeneration. *Investigative Ophthalmology & Visual Science*, 38, 469–477.
- Molloy, M. F., Bahg, G., Li, X., Steyvers, M., Lu, Z.-L., & Turner, B. M. (2018). Hierarchical Bayesian Analyses for Modeling BOLD Time Series Data. *Computational Brain & Behavior*, 1(2), 184–213.
- Molloy, M. F., Bahg, G., Lu, Z.-L., & Turner, B. M. (2019). Individual Differences in the Neural Dynamics of Response Inhibition. *Journal of Cognitive Neuroscience*, 31(12), 1976–1996.
- Nordhausen, K., Sirkia, S., Oja, H., & Tyler, D. E. (2018). Tools for Multivariate Nonparametrics. *Package 'ICSNP' in CRAN repository*. Retrieved from: <https://cran.r-project.org/package=ICSNP>.
- Ou, W. C., Lesmes, L. A., Christie, A. H., Denlar, R. A., & Csaky, K. G. (2021). Normal- and Low-Luminance Automated Quantitative Contrast Sensitivity Assessment in Eyes With Age-Related Macular Degeneration. *American Journal of Ophthalmology*, 226, 148–155.
- Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research*, 23, 689–699.
- Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z.-L., Steyvers, M., & Turner, B. M. (2018). A Tutorial on Joint Models of Neural and Behavioral Measures of Cognition. *Journal of Mathematical Psychology*, 84, 20–48.
- Pesudovs, K., Hazel, C. A., Doran, R. M., & Elliott, D. B. (2004). The usefulness of Vistech and FACT contrast sensitivity charts for cataract and refractive surgery outcomes research. *British Journal of Ophthalmology*, 88, 11–16.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*. Retrieved from: <https://www.r-project.org/nosvn/conferences/DSC-2003/Drafts/Plummer.pdf>.
- Prins, N. (2013). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision*, 13(7), 3.
- R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ramulu, P., Dave, P., & Friedman, D. (2015). Precision of contrast sensitivity testing in glaucoma. *ARVO Annual Meeting Abstracts*, 56, 2225.
- Reynaud, A., Tang, Y., Zhou, Y., & Hess, R. (2014). A unified framework and normative dataset for second-order sensitivity using the quick contrast sensitivity function (qCSF). *Journal of Vision*, 14(10), 1428.
- Reum, J. C. P., Hovel, R. A., & Greene, C. M. (2015). Estimating Continuous Body Size-Based Shifts in Delta N-15-Delta C-13 Space Using Multivariate Hierarchical Models. *Marine Biology*, 162(2), 469–478.
- Rohaly, A. M., & Owsley, C. (1993). Modeling the contrast-sensitivity functions of older adults. *Journal of the Optical Society of America A, Optics and Image Science*, 10, 1591–1599.

- Rosen, R., Jayaraj, J., Bharadwaj, S. R., Weeber, H. A., Van der Mooren, M., & Piers, P. A. (2015). Contrast sensitivity in patients with macular degeneration. *ARVO Annual Meeting Abstracts*, 56, 2224.
- Rosén, R., Lundström, L., Venkataraman, A. P., Winter, S., & Unsbo, P. (2014). Quick contrast sensitivity measurements in the periphery. *Journal of Vision*, 14(8):3, 1–10.
- Rouder, J. N., & Lu, J. (2005). An Introduction to Bayesian Hierarchical Models with an Application in the Theory of Signal Detection. *Psychonomic Bulletin & Review*, 12(4), 573–604.
- Rouder, J. N., Sun, D. C., Speckman, P. L., Lu, J., & Zhou, D. (2003). A Hierarchical Bayesian Statistical Framework for Response Time Distributions. *Psychometrika*, 68(4), 589–606.
- Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, 122, 105–123.
- Schütt, H. H., & Wichmann, F. A. (2017). An image-computable psychophysical spatial vision model. *Journal of Vision*, 17(12), 12.
- Song, M., Behmanesh, I., Moaveni, B., & Papadimitriou, C. (2020). Accounting for Modeling Errors and Inherent Structural Variability through a Hierarchical Bayesian Model Updating Approach: An Overview. *Sensors*, 20(14), 3874.
- Stellmann, J. P., Young, K. L., Pottgen, J., Dorr, M., & Heesen, C. (2015). Introducing a New Method to Assess Vision: Computer-Adaptive Contrast-Sensitivity Testing Predicts Visual Functioning Better than Charts in Multiple Sclerosis Patients. *Multiple Sclerosis Journal - Experimental, Translational & Clinical*, 1, 1–8.
- Storz, J. F., & Beaumont, M. A. (2002). Testing for Genetic Evidence of Population Expansion and Contraction: An Empirical Analysis of Microsatellite DNA Variation Using a Hierarchical Bayesian Model. *Evolution*, 56(1), 154–166.
- Tan, D. T. H., & Fong, A. (2008). Efficacy of neural vision therapy to enhance contrast sensitivity function and visual acuity in low myopia. *Journal of Cataract & Refractive Surgery*, 34, 570–577.
- Thall, P. F., Wathen, J. K., Bekele, B. N., Champlin, R. E., Baker, L. H., & Benjamin, R. S. (2003). Hierarchical Bayesian Approaches to Phase II Trials in Diseases with Multiple Subtypes. *Statistics in Medicine*, 22(5), 763–780.
- Thomas, M., Silverman, R. F., Vingopoulos, F., Kasetty, M., Yu, G., & Kim, E. L. et al. (2021). Active Learning of Contrast Sensitivity to Assess Visual Function in Macula-Off Retinal Detachment. *Journal of VitreoRetinal Diseases*, 5(4), 313–320.
- Thrane, E., & Talbot, C. (2019). An Introduction to Bayesian Inference in Gravitational-Wave Astronomy: Parameter Estimation, Model Selection, and Hierarchical Models. *Publications of the Astronomical Society of Australia* 36, e010.
- Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, 35, 2503–2522.
- Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian Framework for Simultaneously Modeling Neural and Behavioral Data. *NeuroImage*, 72, 193–206.
- van Gaalen, K. W., Jansonius, N. M., Koopmans, S. A., Terwee, T., & Kooijman, A. C. (2009). Relationship between contrast sensitivity and spherical aberration: Comparison of 7 contrast sensitivity tests with natural and artificial pupils in healthy eyes. *Journal of Cataract & Refractive Surgery*, 35, 47–56.
- Vingopoulos, F., Wai, K. M., Katz, R., Vavvas, D. G., Kim, L. A., & Miller, J. B. (2021). Measuring the Contrast Sensitivity Function in Non-Neovascular and Neovascular Age-Related Macular Degeneration: The Quantitative Contrast Sensitivity Function Test. *Journal of Clinical Medicine*, 10(13), 2768.
- Wai, K. M., Vingopoulos, F., Garg, I., Kasetty, M., Silverman, R. F., & Katz, R. et al. (2021). Contrast sensitivity function in patients with macular disease and good visual acuity. *British Journal of Ophthalmology*, <https://doi.org/10.1136/bjophthalmol-2020-318494>. [e-pub ahead of print].
- Wang, C., Lin, X., & Nelson, K. P. (2020). Bayesian Hierarchical Latent Class Models for Estimating Diagnostic Accuracy. *Statistical Methods in Medical Research*, 29(4), 1112–1128.
- Watson, A. B. (2017). QUEST+: A general multi-dimensional Bayesian adaptive psychometric method. *Journal of Vision*, 17(3):10, 1–27.
- Watson, A.B. (2000). Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express*, 6(1), 12–33.
- Watson, A. B., & Ahumada, A. J. (2005). A Standard Model for Foveal Detection of Spatial Contrast. *Journal of Vision*, 5(9), 717–740.
- Watson, A.B., & Malo, J. (2002). Video quality measures based on the standard spatial observer. *Proceedings of the IEEE International Conference on Image Processing*, 3, 41–44.

- Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*, 1293–1313.
- Wikle, C. K. (2003). Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes. *Ecology*, *84*(6), 1382–1394.
- Wilcox, R. (2012). *Modern Statistics for the Social and Behavioral Sciences: A Practical Introduction* (pp. 101–102). Boca Raton, Florida, USA: CRC Press.
- Wilson, J. D., Cranmer, S., & Lu, Z.-L. (2020). A Hierarchical Latent Space Network Model for Population Studies of Functional Connectivity. *Computational Brain & Behavior*, *3*, 384–399.
- Yan, F.-F., Hou, F., Lu, Z.-L., Hu, X., & Huang, C.-B. (2017). Efficient Characterization and Classification of Contrast Sensitivity Functions in Aging. *Scientific Reports*, *7*, 5045.
- Yang, J., Zhu, H., Choi, T., & Cox, D. D. (2016). Smoothing and Mean-Covariance Estimation of Functional Data with a Bayesian Hierarchical Model. *Bayesian Analysis*, *11*(3), 649–670.
- Zhao, Y., Lesmes, L. A., Dorr, M., & Lu, Z.-L. (2021). Quantifying Uncertainty of the Estimated Visual Acuity Behavioral Function With Hierarchical Bayesian Modeling. *Translational Vision Science & Technology*, *10*(12), 18.
- Zhou, Y. F., Huang, C.-B., Xu, P. J., Tao, L. M., Qiu, Z. P., & Li, X. R. et al. (2006). Perceptual learning improves contrast sensitivity and visual acuity in adults with anisometric amblyopia. *Vision Research*, *46*(5), 739–750.