Characterization of Field Loss Based on Microperimetry Is Predictive of Face Recognition Difficulties

Thomas S. A. Wallis,^{1,2} Christopher Patrick Taylor,¹ Jennifer Wallis,³ Mary Lou Jackson,³ and Peter J. Bex¹

¹Schepens Eye Research Institute, Harvard Medical School, Boston, Massachusetts

²School of Psychology, The University of Western Australia, Perth, Australia

³Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts

Correspondence: Thomas S. A. Wallis, Centre for Integrative Neuroscience and Department of Computer Science, University of Tübingen, Tübingen, Germany; thomas.wallis@uni-tuebingen.de.

TSAW, CPT, and JW contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Submitted: May 16, 2013

Accepted: November 4, 2013

Citation: Wallis TSA, Taylor CP, Wallis J, Jackson ML, Bex PJ. Characterization of field loss based on microperimetry is predictive of face recognition difficulties. *Invest Ophthalmol Vis Sci*. 2014;55:142–153. DOI:10.1167/iov.13-12420

PURPOSE. To determine how visual field loss as assessed by microperimetry is correlated with deficits in face recognition.

METHODS. Twelve patients (age range, 26–70 years) with impaired visual sensitivity in the central visual field caused by a variety of pathologies and 12 normally sighted controls (control subject [CS] group; age range, 20–68 years) performed a face recognition task for blurred and unblurred faces. For patients, we assessed central visual field loss using microperimetry, fixation stability, Pelli-Robson contrast sensitivity, and letter acuity.

RESULTS. Patients were divided into two groups by microperimetry: a low vision (LV) group ($n = 8$) had impaired sensitivity at the anatomical fovea and/or poor fixation stability, whereas a low vision that excluded the fovea (LV:F) group ($n = 4$) was characterized by at least some residual foveal sensitivity but insensitivity in other retinal regions. The LV group performed worse than the other groups at all blur levels, whereas the performance of the LV:F group was not credibly different from that of the CS group. The performance of the CS and LV:F groups deteriorated as blur increased, whereas the LV group showed consistently poor performance regardless of blur. Visual acuity and fixation stability were correlated with face recognition performance.

CONCLUSIONS. Persons diagnosed as having disease affecting the central visual field can recognize faces as well as persons with no visual disease provided that they have residual sensitivity in the anatomical fovea and show stable fixation patterns. Performance in this task is limited by the upper resolution of nonfoveal vision or image blur, whichever is worse.

Keywords: face recognition, central vision loss, low vision, perimetry, Bayesian inference

A crucial function for independent living is recognizing people around us. Facial features such as the eyes¹ are the primary visual cues most humans use in discriminating people they know from those they do not know. Reduced sensitivity in the central visual field (defined here as the central 20° of vision) presents a set of challenges for patients to manage as part of their daily lives. For observers with central visual field loss, face perception can be greatly impaired, and individuals who present for vision rehabilitation regularly report difficulty with face perception.^{2,3} Because photoreceptor and ganglion cell density,^{4,5} as well as contrast sensitivity,^{6–10} decrease as retinal eccentricity increases, it is no surprise that patients with sensitivity losses in the central visual field have difficulty perceiving faces (see Kwon and Legge¹¹ for a quantification of the human contrast sensitivity function contribution to face recognition in the fovea and periphery).

Numerous studies have found impaired face recognition^{12–18} and emotional expression perception^{19,20} in individuals with central visual field loss. For example, Bullimore et al.¹² determined a threshold viewing distance to reach a criterion level on tasks of identity and expression recognition. Patients with central field loss performed worse (required a nearer viewing distance) than normally sighted controls and were generally better at recognizing the expression of a face

than its identity. In addition, word reading acuity was better correlated with face recognition thresholds than letter chart acuity, grating acuity, and contrast threshold for edge detection. In a face recognition task similar to the one we use here, patients with AMD showed slower reaction times and poorer face recognition performance than age-matched controls.¹³ Tejeria et al.¹⁷ correlated measures of visual function (letter acuity, reading, Pelli-Robson contrast sensitivity, and color vision) and self-reported difficulty in face recognition with performance on a familiar face identification and an “odd one out” expression discrimination task in 30 subjects with AMD. Letter acuity and reading acuity were found to be correlated with performance on the face perception tasks, but there was little evidence for a relationship between self-reported difficulty in face perception and performance in either of the tasks. Using a “celebrity”/“not a celebrity” face recognition task, Peli et al.¹⁶ found that certain types of digital image enhancement improved recognition performance in some (16 of 38) patients with central vision loss (see also Peli et al.²¹). Patients with central field loss also demonstrated poorer performance on a famous face recognition task relative to controls in a study by Dulin et al.²² It is worth noting, however, that in face recognition tasks relying on famous faces the contribution of

cultural and general visual exposure cannot be separated from face identification performance per se.

Many of these studies have examined correlations between face recognition performance and standard metrics of functional vision such as acuity and contrast sensitivity and found modest to strong relationships depending on the type of stimulus used in assessment. However, the relationships between these quantities and the extent of central field loss are uncertain. To more precisely measure these relationships, we combine in the present study an objective measure of face recognition with clinical characterization of retinal deficiencies using fundus-oriented microperimetry. This face identification task has been used previously²³ and shows many classic effects from the face perception literature, including the inversion effect,^{23,24} the contrast-reversed face effect,²⁵ the preferential use of horizontal orientation information,^{26–28} and spatial frequency tuning of approximately 10 cycles per face.²³ In addition to letter acuity and contrast sensitivity, visual function was assessed with fundus-oriented microperimetry, which allows precise assessment of light sensitivity over a large (approximately 20° diameter) region of the retina. Perimetry also measures each patient's fixation stability and the locus of fixation, which may be foveal or nonfoveal.

Individuals with foveal vision loss may use nonfoveal areas of the retina for face perception.²⁹ The peripheral retina has reduced sensitivity to high spatial frequencies compared with the fovea. We varied the amount of blur present in our face stimuli by selectively removing high spatial frequency content from the images. Individuals with foveal vision loss should be relatively unaffected by stimulus blur if blur only affects the high spatial frequencies to which they are insensitive. We classified low vision participants into two groups on the basis of their sensitivity across the central visual field, as well as the stability of their fixation in microperimetry. One group had impaired foveal vision and/or poor fixation stability (LV group), whereas the other group had visual impairments that excluded the fovea (LV:F group). We hypothesized that the group with intact foveal function would show a consistent reduction in face recognition as a function of blur, whereas the impaired fovea group would show poor performance for even relatively blur-free stimuli consistent with the use of peripheral vision.

METHODS

Participants

Patients with various pathologies affecting central visual field sensitivity were recruited from the Vision Rehabilitation Clinic at the Massachusetts Eye and Ear Infirmary. Twenty-one patients with low vision were initially enrolled in the study. Eight of these patients chose to withdraw after completing less than one full block of trials, in all cases citing frustration with task difficulty. This introduced a self-selection bias that may have led us to overestimate the performance of patients with low vision in our study. One additional patient was excluded from analyses because we did not have accompanying microperimetry data for that patient. These exclusions left a final group of 12 patients aged 26 to 70 years (mean age, 48.8 years). Three of the final group of patients completed less than the full three blocks because of time constraints (but completed at least one block). These were P07 (one block), P08 (two blocks), and P09 (one block). All patients were tested after initial consultation at the clinic with an ophthalmologist (MLJ), but before any vision rehabilitation training with an occupational therapist.

Twelve controls with normal retinal function were recruited, including two of the authors (C01 and C09). These subjects

ranged in age from 20 to 68 years (mean age, 40.8 years). Demographic information for all subjects is summarized in Table 1. All subjects provided informed consent; the procedures of the experiment were approved by the Institutional Review Board of the Massachusetts Eye and Ear Infirmary, and the experiment was conducted in accord with the Declaration of Helsinki.

All participants were divided into one of three groups for data analysis. Control subjects (CS group) were those with no known visual pathology. Patients with low vision were divided into the following two subgroups based on clinical evaluations and perimetry: those with at least partially functional foveal vision (LV:F group) and those without (LV group) (Fig. 1). This classification was made for the eye with better acuity or for the eye with more stable fixation in microperimetry if acuity was equal in the two eyes. The testing area for microperimetry was determined by initial fixation, so it was not centered around the anatomical fovea for all patients. However, the testing array always included the anatomical fovea. The area of the anatomical fovea was defined as the region of the retina located two vertical optic disk diameters to the temporal side of the optic disk.³⁰ Three conditions had to be met for a patient to be classified in the LV:F group. First, the perimetry array had to be centered around the anatomical fovea. Second, the patient had to report seeing at least one of the four perimetry targets in the center of the array. Third, the patient had to show stable fixation (i.e., >95% of fixations fell within 2° of the anatomical fovea). Other patients with a low vision diagnosis failed to detect perimetry targets presented at the fovea and had unstable fixation (LV group). All patients were classified using these criteria by one of us (JW). The classification was independently assessed by familiarizing an occupational therapist of the Vision Rehabilitation Clinic (not an author) with the classification guidelines; this person made the same classifications as JW.

Apparatus and Stimuli

Face stimuli were presented on an iMac (late 2011 model; Apple, Inc., Cupertino, CA) with a 27-inch LCD monitor with a spatial resolution of 2560 × 1440 pixels and a temporal resolution of 60 Hz using MATLAB (MathWorks, Natick, MA) and Psychtoolbox-3 (provided in the public domain by www.psychtoolbox.org).^{31,32} The observers viewed the display under normal room illumination, and the mean luminance of the display was 250 cd/m². To increase the contrast resolution of the display, a bit-stealing routine was used to increase the number of grey levels from 6 to 8.8 bits.³³ The stimuli were monochromatic photographs of faces presented at three levels of blur. Figure 2 shows examples of representative face stimuli. A detailed description of how the faces were photographed and processed is provided by Gaspar et al.²³; thus, only a brief description will be provided here. The face stimuli used in the experiment are available upon request from the authors. The face set contains photographs of 10 individuals (five female), all Caucasian, with a mean age of 24 years; none of the models had facial hair, visible piercings, tattoos, or eyeglasses. All the face images were cropped with an oval window using a ratio of width to height of 1.5 to remove all clothing and hair. Each model was photographed from five viewpoints. To produce each viewpoint image, the model was instructed to fixate on a marker positioned on the wall behind the camera. The face set used was made up of two left-facing, two right-facing, and one front-facing view of each model, for a total of 50 face images. To control for possible variations in detectability among the face images,³⁴ we equated the Fourier amplitude spectra of the images by computing the average amplitude spectrum of the set of face images and then replacing the amplitude spectrum

TABLE 1. Demographic and Clinical Information

Patient or Control	Age, y	Sex	Instruction	PR OD	PR OS	VA OD	VA OS	LogMAR VA in Better Eye	Diagnosis
P01	52	M	No	NA	1.5	NA	20/150–200	0.88	Right prosthesis, left keratoplasty
P02	57	M	Yes	1.5 (Binocular)	...	20/400	20/150–1	0.88	Congenital nystagmus
P03	34	M	No	1.35	1.35	20/40	20/30	0.18	Retinitis pigmentosa
P04	26	F	Yes	1.65 (Binocular)	...	20/150	20/150	0.88	Stargardt dystrophy
P05	30	M	Yes	1.05 (Binocular)	...	1/12	1/16	1.08	Macular malfunction
P06	57	F	No	1.2	0.45	20/30	20/200	0.18	Exposure keratopathy/neurotrophic cornea in left eye
P07	70	F	No	1.05 (Binocular)	...	20/100	20/150	0.70	Glaucoma
P08	32	M	No	1.65 (Binocular)	...	20/150	20/150	0.88	Achromatopsia
P09	70	M	No	0.45 (Binocular)	...	20/400	20/200	1.00	Anterior ischemic optic neuropathy
P10	69	F	No	1.35 (Binocular)	...	20/70	20/80	0.54	AMD
P11	58	F	No	1.05 (Binocular)	...	20/70	20/100	0.54	Retinal degeneration
P12	31	F	No	1.65	1.65	20/25	20/20	0.00	Decreased vision after trauma
C01	29	M	...	1.95	...	...	...	−0.13	...
C02	20	F	...	1.95	...	...	...	−0.13	...
C03	46	F	...	1.8	...	...	...	0.18	...
C04	68	F	...	NA	...	...	...	NA	...
C05	40	F	...	NA	...	...	...	NA	...
C06	60	F	...	NA	...	...	...	NA	...
C07	61	F	...	1.65	...	...	...	NA	...
C08	34	M	...	1.8	...	...	...	NA	...
C09	45	M	...	1.8	...	...	...	0.00	...
C10	23	F	...	1.8	...	...	...	−0.13	...
C11	34	M	...	1.95	...	...	...	−0.13	...
C12	29	F	...	1.8	...	...	...	−0.13	...

The “Instruction” column refers to those patients who received the pilot instruction manipulation. Visual acuity (VA) and Pelli-Robson contrast sensitivity (PR) were measured binocularly in controls and where marked in patients. Ellipsis, not applicable; logMAR, logarithm of the minimum angle of resolution; NA, not available.

of each image with the average amplitude spectrum. As noted in the introduction, this stimulus set has been previously shown to demonstrate classic effects from the face perception literature. This indicates that our stimuli tap mechanisms similar to those used in more complex stimulus sets, despite

using only a relatively small (10 individuals) and relatively unnatural (lacking features such as hair and skin pigment) set of faces.

To measure the effect of blur on recognition performance of both patients and controls, we tested face recognition

FIGURE 1. Classification of patients into groups based on microperimetry. (A) Schematic of classification. First, the location of the anatomical fovea (marked here by a red “F”) was estimated by measuring two vertical optic disk diameters (red line) to the temporal side of the optic disk. If any of the targets around the anatomical fovea were detected and greater than 95% of fixations (blue crosses) were within 2° of the anatomical fovea in the eye with greater acuity, the patient was classified into the LV:F group. (B) Microperimetry output for a patient in the present study (P04) for the patient’s right eye (OD), tested with a modified threshold strategy. The foveal area is obscured by a scotoma. The patient attempted to fixate in the foveal region. Fixation stability is poor (lower polar plot, blue crosses), and no targets presented to the anatomical fovea were detected (red circles).

FIGURE 2. Stimuli used in the experiment (by Gaspar et al.,²³ used with permission). *Left:* Each stimulus face; each column shows one of the five possible viewpoints. *Right:* An example of the blur kernels used. The *top image* depicts the cutoff of 32 cycles per image; the *bottom image* depicts the cutoff of 16 cycles per image.

performance at two levels of synthetic blur and a no-blur condition. Blurred images were produced by applying a low-pass ideal filter (high-frequency cutoff of 32 or 16 cycles per image, approximately 8 or 4 cycles per degree) to the Fourier amplitude spectrum of each image (Fig. 2). A zero mean Gaussian pixel noise (root mean square contrast, $\sigma = 0.005$) was added to the test image on each trial to allow comparison with the ideal observer (see Supplementary Fig. S1).

Perimetry

Microperimetry testing was conducted with an optical coherence tomography/scanning laser ophthalmoscopy microperimeter (Optos plc, Dunfermline, Scotland). The extent of central field loss was assessed by having patients fixate a cross presented in the center of the display, while a Goldmann III test stimulus (0.43° disk) was presented for 200 milliseconds (ms) at each of 52 radially organized locations within a 21° -diameter circle centered around the initial fixation location. Seven patients (P03, P04, P06, P07, P08, P11, and P12) were tested with a modified threshold strategy, and five patients (P01, P02, P05, P09, and P10) were tested with a suprathreshold strategy. The testing regimen was chosen depending on the clinical appropriateness for each patient, not as part of a research protocol. Suprathreshold testing entails the presentation of a target at an intensity that is anticipated to be seen (0-decibel [dB] attenuation), and the patient indicates when the target is detected by pressing a button. In modified stepwise threshold testing (using a 0-10-16 technique), a -10 -dB point is presented on the first trial and then lowered to -16 dB if seen or to 0 dB if not seen. Testing positions that elicit no response suggest field loss. During testing, fixation is continuously monitored. Thus, a measure of fixation stability can be obtained (defined as fixations that remain within 2° and 4° of initial fixation). Unfortunately, we were unable to save raw fixation coordinates because of computer malfunction, meaning that the more sensitive fixation measure of the bivariate contour ellipse area (BCEA) (e.g., see Crossland et al.³⁵) could not be calculated. However, because we note credible correlations with our relatively blunt measure of fixation stability, the BCEA would likely only strengthen the relationships we observe.

All but one patient failed to report at least one microperimetry target entirely. The remaining patient (P06), tested in a modified threshold regimen, detected all targets but required increased brightness relative to unimpaired vision (relative deficits). All other patients who were tested with a modified threshold strategy showed relative deficits in addition to their failures to report targets at the highest brightness (e.g.,

P04 in Fig. 1). Therefore, all patients can be considered to have had some central vision loss (Table 2).

Procedure

Participants completed a single-interval 10-alternative forced-choice face identification task. Observers freely viewed the display (binocular presentation, with no restriction on eye movements) from approximately 57 cm, and at this distance the test face images subtended $8.4^\circ \times 8.8^\circ$ (width $\times$ height) of visual angle. On each trial, the observer saw a fixation point in the center of the screen presented for 200 ms, followed by a test face image presented in the center of the screen for 500 ms. The identity, viewpoint, and blur level of the test face was randomly selected at each trial. After the test image disappeared, 10 smaller images ($6.1^\circ \times 6.4^\circ$ of visual angle), one of each face, were presented in a grid of five columns by two rows on the screen. These faces were of the 10 individuals from the face set and were presented in random order at the same viewpoint as the test image but without blur. The observer used the arrow keys on the keyboard to move a response selection rectangle over the response faces and then

TABLE 2. Perimetry Data for Patients With Low Vision

Patient Code	2° Fixation, %	4° Fixation, %	Perimetry Targets Missed, %	Analysis Group
P01	57	94	2	LV
P02	9	37	46	LV
P03	100	100	67	LV:F
P04	41	99	13	LV
P05	42	100	44	LV
P06	100	100	0	LV:F
P07	18	100	42	LV
P08	41	80	2	LV
P09	76	100	83	LV
P10	100	100	6	LV:F
P11	72	100	4	LV
P12	100	100	2	LV:F

The fixation percentages denote the percentage of all fixations made within 2° or 4° of the center of the testing grid. Any subject who both showed good fixation stability and reported seeing any of the four perimetry targets presented around the anatomical fovea was classified as having a functional fovea (LV:F group). While P06 did not miss any perimetry targets at the highest intensity, extensive relative deficits were evident with lower-intensity targets.

pressed the space bar to enter the response. If the observer was unable to use the keyboard, he or she pointed at a face, and the experimenter keyed in the response. The 10 response selection faces remained on-screen until a response was entered. If the observer's response was correct, then a clearly audible high-pitched tone was played; if the observer's response was incorrect, a tone of lower pitch was played. Before the experimental trials, all observers completed five practice trials to introduce them to the task.

Each observer completed three blocks of 75 trials, except for one control (C02), who completed three blocks of 150 trials, and three patients (P07, P08, and P09), who completed only one, two, and one blocks, respectively. The three levels of blur were each tested 25 times per block in a random order for each subject.

Three of the patients with low vision (P02, P04, and P05) received an instruction manipulation designed to improve their guidance of eye movements to important regions of the faces. After completing the first block of trials, these subjects were instructed to "focus on the eyes, nose, and mouth" by the experimenter, who asked the subject to practice by looking at the experimenter's face. This manipulation was designed to ensure that patients were placing regions of functional retina over facial features important for identification as far as possible. However, a pilot analysis showed no evidence that the instruction improved performance beyond a simple learning effect. All of these patients reported that they were already able to place their gaze to perceive important facial features, making it unlikely that this instruction manipulation would have had any further effect. Therefore, we group together patients who received this instruction manipulation with patients who did not in further analyses.

Data Analysis

To analyze these data, we used a multilevel logistic generalized linear model (see Supplementary Materials Section 1.2 for details). Multilevel models are extensions of traditional regression models that treat regression parameters themselves as conditional on higher-order structure in the data.^{36,37} Our data have hierarchical structure in that some parameters (e.g., a given subject's performance level) can be considered to arise from distributions of mean performance within a group (e.g., "control" or "low vision"), which themselves are conditional upon an overlying population distribution of mean performance level (the distribution for humans). Intuitively, this is similar to running three regressions: one on the data for each subject to find the parameter weights at the subject level, a second on the weights from the first regression to estimate the group parameters, and a third on the group estimates to find the population-level parameters. However, all of these parameters are estimated simultaneously rather than in a stepwise fashion as implied by this example.

A multilevel model like the one we use is a more general form of the more common multilevel generalized linear mixed model, which has been recently promoted within the psychophysics^{38,39} and clinical vision⁴⁰ literature because it allows the analysis of data on the population level. A mixed model is "mixed" because it contains both fixed effects (in which a parameter is not allowed to vary at some level of the model) and random effects (for which the uncertainty in the parameter estimates is considered throughout all levels of the model). One way to consider a fixed effect is that it is a special case of a random effect in which the variance parameter is set to be zero (i.e., there is no uncertainty around the point estimate of the parameter); see Gelman and Hill.^{37(p245)} In the framework we use, all parameters can be thought of as random effects.³⁷ This treatment has the desirable characteristics that

uncertainty at all levels of the model is considered at all others (rather than reducing the model to a point estimate at some level) and that this uncertainty can be propagated into predictions. For our data set, this is particularly useful because not all subjects completed the same number of trials. Therefore, our approach takes the individual subject uncertainties into account when we make inferences at the group level.

We base statistical inferences on the full Bayesian posterior distribution of credible parameter values rather than on null hypothesis significance testing. Desirable properties of this approach are considered briefly in the Discussion section. Because the posterior distributions for the models we use are not derivable analytically, we numerically estimate proportional distributions using Markov Chain Monte Carlo (MCMC) sampling implemented in the Stan package (<http://mc-stan.org/> [in the public domain]).^{41,42} Results were analyzed with the R statistical computing environment (R Foundation for Statistical Computing, Vienna, Austria)⁴³ using additional packages.^{44–46} The estimate of the posterior distribution for the primary analysis is based on four independent chains of 100,000 samples using the No-U-Turn Sampler II (<http://mc-stan.org/> [in the public domain]) sampling method,⁴² with half of those samples being warm-up (adapting the step size of the sampler) and saving every 100th sample to reduce both autocorrelation in the chains and data file size, resulting in 2000 final samples. Convergence was assessed by the $\hat{R}$ value (the ratio of between to within chain variance) (see Gelman and Rubin⁴⁷), as well as by inspection of traceplots. Code to perform these analyses and raw data can be found on the first author's GitHub page (http://github.com/tomwallis/microperimetry_faces [in the public domain]).

We were primarily interested in how performance differed across the three groups as a function of the blur level. In addition, because there could have been a learning effect and not all participants completed the same number of blocks, we included a covariate for block. The trial by trial (correct or incorrect) data for each subject were modeled as a linear function of the log high-frequency cutoff (*HFC*) of the blur filter and the block as follows:

$$\eta = a + b \times \log_{10}(HFC) + c \times \text{block} \quad (1)$$

This yielded the following three free parameters for each subject: intercept a , blur slope b , and block slope (linear learning effect) c . Including age as an additional covariate did not substantively change our conclusions, and other covariates such as visual acuity cannot be included in this model because they were reported as corrected to normal for controls, in which case there is no variation in acuity among controls. To make the regression coefficients easier to interpret, we centered each input predictor by subtracting the mean (or two in the case of block); see Gelman and Hill.^{37(p55)} Therefore the intercept parameter represents performance when blur was at its mean in the second block. We additionally examined a model (data not shown) that contained an interaction term between block and blur slope. The interaction term was not credibly different from zero, so we excluded it for simplicity. The linear predictor η is then passed through the inverse logit function to give the expected value for a Bernoulli trial, with an additional lower bound of 0.1 (reflecting chance performance).

For the two conditions in which we imposed physical stimulus blur, the high-frequency cutoffs correspond to 32 cycles per image (less blurry) and 16 cycles per image (more blurry). However, no frequencies were filtered for the unblurred condition. To express the unblurred condition on the same scale, we assume the high-frequency cutoff to be the

FIGURE 3. Face recognition performance as a function of high-frequency cutoff (blur level) and block for each subject (note the logarithmic x -axis). A lower cutoff value corresponds to a more blurred stimulus. The subject code displayed in the *top margin* corresponds to Table 1, and the *side text* shows the analysis group for that subject. *Points* show the average performance of the subject; *error bars on points* denote 95% beta distribution confidence intervals. *Solid lines* represent the predictions of the Bayesian multilevel model at the subject level, and *shaded regions* represent 95% credible intervals for these predictions.

Nyquist limit of the stimulus, which given that our images were 372 pixels squared the highest frequency present in our stimuli was 186 cycles per image, approximately 46 cycles per degree.

Each coefficient from each subject is hypothesized to arise from a distribution for that subject's group. A population hyperparameter is also placed over all groups, causing group differences to be pulled toward the population mean and making our inferences on group differences more conservative where parameter estimates are uncertain. The full specification of the model is provided in Supplementary Materials Section 1.2, along with a discussion of the tolerance of our conclusions to the prior distributions chosen. All groups are given the same mean (zero) and variance (broad) in their regression parameters a priori, so it is not the case that our use of priors trivially presumes the conclusion we report.

RESULTS

Recognition Performance

Figure 3 shows recognition performance for each subject for each blur level and block, with the corresponding fits of the multilevel model at the subject level. Each MCMC sample after warm-up represents a plausible value of the posterior probability over the full parameter space. We computed a distribution of expected proportion correct for a continuum of blur values from the subject-level parameters a , b , and c for each of the post-warm-up samples of the MCMC chains after pooling the chains. The solid curves in Figure 3 show the mean of this distribution for each subject, and the shaded regions are the values containing 95% of the distribution's mass, the highest-density interval (HDI). Unlike a frequentist confidence interval, the 95% HDI of a posterior distribution allows us to

FIGURE 4. Face recognition performance as a function of high-frequency cutoff (blur level) for each group (note the log unit x-axis). A lower cutoff value corresponds to a more blurred stimulus. Points show each group mean, and error bars denote $\pm$ SEM between subjects. Solid black lines represent the predictions of the Bayesian multilevel model at the group level, and shaded regions represent 95% credible intervals on these predictions. Faint grey lines connect the mean proportion correct for each subject (data are replotted from Fig. 3 and averaged across block).

say that given the model and the data there is a 95% chance that the true parameter value lies in this interval. These are 95% credible intervals (as distinct from confidence intervals⁴⁸). As expected, performance increased as a function of the logarithm of high-frequency cutoff (i.e., decreased as a function of blur level) for most subjects. The model provides reasonable fits to the data at the individual subject level.

Each subject's three free parameters are conditional on the parameters of their group, which are themselves conditional on the population-level parameters. Therefore, the data from one subject inform the fits to the data of another subject, more so for subjects in the same group. Statistical inferences at the group level can be made by comparing the group-level parameter distributions. Of principal interest here is the group-level distribution for performance as a function of blur.

Predictions at the group level are shown in Figure 4 and were derived from the MCMC samples as in the subject-level data. It can be seen that the controls (CS group) had higher performance than the patients with low vision (LV group) using diseased retina or with unstable fixation. In addition, those patients with low vision with remaining foveal sensitivity (LV:F group) are similar to controls in their performance.

The group differences shown in Figure 4 can be formally compared by calculating the difference score between parameter estimates in each MCMC sample to create a distribution of difference scores. That is, we calculate the value of the linear predictor for each MCMC sample (see Equation 1) at each block and blur level; we then apply the inverse logit to convert these values into predicted probability correct and then take a difference score between the groups for each sample. We then base our inferences about whether two groups have credibly different performance on the distribution of difference scores between those groups by asking whether zero (i.e., no difference) is within the range of credible values.⁴⁸ Note that no correction is necessary for making multiple comparisons because all possible comparisons are encapsulated by the full posterior distribution,⁴⁹ which remains unchanged whether we compare two parameters or many. The posterior distributions for the group-level parameters are provided in Supplementary Figure S2.

FIGURE 5. Means and HDIs of pairwise comparisons between groups. These values are calculated from the posterior distributions of difference scores in predicted proportion correct. Differences are shown for each blur level at each block of trials. At all blur levels and as early as block one, the CS and LV:F groups performed better than the LV group.

At all blur levels tested, and as early as the first 25 trials, the performance of the CS group was credibly higher than that of the LV group (Fig. 5). In addition, the LV group was credibly lower than the LV:F group. No credible differences were evident between the CS and LV:F groups. Note that in this model, the values for parameters a , b , and c are estimated from all the data. While strictly this means that the parameters are not independent of block, this is not important for our statement that group differences emerge as early as block one. The same group differences are evident after fitting a two parameter (a and b) model to only the data from block one.

We can also examine both the absolute slope parameters for blur and block and differences between groups for these slope parameters. Slopes credibly greater than zero signify a linear relationship between (the logit of) performance and the variable of interest (log high-frequency cutoff for b and block for c [see Equation 1]). The specific value of the slope parameter depends on the scaling of the predictor because the slope multiplies the value of the input predictor. If the value of the parameter is zero, this indicates no linear relationship (i.e., the logit of performance does not change linearly as a function of the predictor).

For the slope as a function of the log high-frequency cutoff (parameter b), performance for the CS and LV:F groups increased as high-frequency cutoff increased (mean values, 1.66 [95% HDI, 1.26–2.03] for the CS group and 1.34 [95% HDI, 0.26–2.46] for the LV:F group). That is, performance got worse as the amount of blur increased. However, the slope for the LV group was not credibly different from zero (mean, 0.83 [95% HDI, –0.34 to 1.89]). This indicates that the performance of the LV group was not credibly affected by changes in blur, whereas the other groups showed effects of blur such that performance improved as blur decreased (i.e., as high-frequency cutoff increased). However, these comparisons should be treated with caution because there were also no credible differences between the group slopes (see Supplementary Materials Section 1.3), indicating that there is large uncertainty for these comparisons. While all groups showed a credible learning effect (parameter c was credibly greater than zero [see Supplementary Fig. S2]), there were no credible differences between group learning rates (Supplementary Fig. S3).

FIGURE 6. Scatterplots of face identification performance (proportion correct) in low vision participants as a function of scotoma level, fixation stability within 2°, contrast sensitivity, and letter acuity. The LV group is shown as *white squares*; the LV:F group is shown as *grey diamonds*. Points show the mean performance for each subject, and *error bars* denote beta distribution 95% confidence intervals. Similar patterns were found when performance was broken down for each blur level.

Correlations Between Visual Function Measures

Scatterplots comparing clinical measures of visual function and face identification performance are shown in Figure 6 for the patients with low vision. We estimated the posterior distribution over the correlation coefficients between variables using MCMC as in the main analysis; these are summarized in Table 3. Because these calculations are based on only 12 data points, we urge caution in drawing strong conclusions from the correlations. To encourage conservative inferences, we place a weakly informative prior distribution over the correlation

matrix with greater mass on the unit diagonal. That is, we a priori assign higher probability to correlations around zero (see Supplementary Materials Section 1.2.1) but with high uncertainty.

Using the conservative Spearman rank order ρ , fixation stability and visual acuity were credibly correlated with face recognition performance (i.e., the credible intervals on these comparisons did not overlap zero). More stable fixation and better acuity were associated with better performance. Credible intervals for all other correlation coefficients overlapped zero. Similar results were found using a bootstrapping

TABLE 3. Rank Order Correlations (Spearman ρ) Between Visual Function Measures for Patients With Low Vision Estimated Using a Bayesian Framework

Variable	TM	FS	CS	VA	PC
TM	1.00	−0.11 (−0.48 to 0.29)	−0.26 (−0.62 to 0.13)	0.23 (−0.17 to 0.60)	−0.30 (−0.67 to 0.10)
FS		1.00	−0.14 (−0.50 to 0.23)	−0.37 (−0.71 to 0.03)	0.48 (0.10 to 0.78)
CS			1.00	−0.09 (−0.46 to 0.28)	0.35 (−0.04 to 0.68)
VA				1.00	−0.47 (−0.81 to −0.09)
PC					1.00

The posterior mean rank order correlation coefficient is given in each cell, with 95% credible intervals. CS, Pelli-Robson contrast sensitivity; FS, fixation stability (percentage of fixations within 2°); PC, proportion correct on the face identification task; TM, proportion of targets missed in perimetry; VA, visual acuity. Correlation coefficients that were credibly different from zero are shown in bold.

approach and traditional calculations of correlation coefficients, except that fixation stability and visual acuity were significantly correlated with each other. We believe that our more conservative approach here is warranted given the small number of data points.

DISCUSSION

Many patients with pathologies affecting the fovea are forced to use nonfoveal retina to accomplish visual tasks. We examined how the extent of central field loss affected face identification. The LV group in our study contained individuals who failed to perceive perimetry targets centered over the anatomical fovea at maximum intensity (i.e., had dense central scotomas), demonstrated poor fixation stability, or presented a combination of these deficits (Table 2). This group performed worse in the face identification task compared with the other groups (Figs. 4, 5). Patients classified into the LV:F group saw at least one of the perimetry targets centered over the anatomical fovea and showed stable fixation. Unlike the LV group, the LV:F group performed similarly to controls, despite having either dense or relative deficits in their central visual field (Table 2). The data also suggest that a patient's letter acuity and fixation stability during perimetry are correlated with difficulties in face recognition. For this task, functional macular retina seems to be the primary determinant of performance among a group of patients with varied degrees of paracentral field loss.

We also examined how stimulus blur affected face identification in these groups. We expected that blur would only affect performance in participants who could resolve high spatial frequencies; thus, visual function would be limited by the lower bound of blur in the stimulus or visual resolution. Face identification performance steadily decreased as blur increased for normally sighted individuals and patients having low vision with (at least partially) functional foveal vision: the slope parameters for these groups were credibly different from zero. Patients with low vision and scotomas that included their fovea (LV group) were relatively unaffected by stimulus blur, performing worse than controls and LV:F subjects at every blur level tested and as early as the first block of trials (Fig. 5). As expected, the LV group showed poor performance at all levels of blur, whereas the group with intact foveal function (LV:F group) showed a consistent reduction in performance as a function of blur similar to controls. While we did not track the gaze position of our participants during the face recognition task, the result is consistent with LV subjects using peripheral vision to perform the task. The limiting factor on the performance of the LV group is likely to be the poor resolution of nonfoveal vision rather than stimulus blur at the levels we imposed.

To provide an illustration of the level of impairment demonstrated by LV patients in the context of our experiment, we calculate how much the stimulus would need to be blurred for controls to produce the same performance as the LV group with no stimulus blur. For an unblurred stimulus, patients in the LV group have an average performance level of approximately 46% (Fig. 4). According to the model, the control group would need a blur cutoff of approximately 0.8 cycles per degree to reach a performance of 46% correct on average. A face stimulus filtered with this blur level is shown in Figure 7. Within the context of stimulus blur, the amount of information that LV patients can extract from a face is clearly limited. Of course, many factors influence performance as stimuli move away from the foveal retina (e.g., reduced contrast sensitivity⁹ and crowding⁵⁰). We do not claim that the reduced perfor-

FIGURE 7. *Left:* An unblurred face from the stimulus set. *Right:* The same face blurred to the level our model predicts would produce performance in controls equivalent to that of the LV group with no stimulus blur (see Fig. 4). The right image gives some intuition as to the level of impairment demonstrated by low vision patients in our task.

mance of the LV group here is caused by intrinsic blur alone, nor does this demonstration mean that stimuli appear blurred to LV patients owing to sharpness overconstancy.⁵¹ We provide this illustration to give the reader some intuition as to the level of impairment shown by these patients compared with normally sighted individuals.

Some patients spontaneously used a preferred retinal locus (PRL) during perimetry. We cannot say whether these patients also used a PRL to perform face recognition because gaze position was not monitored in this task. Because these patients showed no sensitivity in the fovea, they were included in the LV group for analysis. In addition, because faces in the recognition task were viewed binocularly and perimetry was monocular, any conclusions we could draw about the influence of a monocular PRL in our task are limited. This may be improved in future studies by using binocular perimetry (currently in development) (see Timberlake G, et al. *IOVS* 2013: ARVO E-Abstract 2182).

Visual acuity and fixation stability correlated with face recognition performance. More stable fixation and better letter acuity were associated with higher task performance. Using our Bayesian estimation of correlation coefficients, we also found no credible association between visual acuity and fixation stability. However, a bootstrapping estimation procedure showed this association to be significant. This contrasts with previous studies^{52–55} that reported no association between visual acuity and fixation stability. The difference can be attributed to our intact foveal function group, who showed stable fixation and relatively good acuity, driving up the strength of this association relative to investigations in which patient populations did not include patients with functional foveal vision. Indeed, any study such as ours that includes diverse groups of patients spanning a wide range of acuities and fixation stabilities is likely to find a stronger correlation than a study within a single group of patients. A scatterplot of fixation stability versus visual acuity (see Supplementary Fig. S4) shows two clusters, one for each of the patient groups, with little evidence of correlations within each group (however, note that the LV:F group was selected on the basis of steady fixation). Of course, these correlational analyses must all be treated cautiously because of our small patient sample size.

It is possible that fixation stability is a more useful correlate of face recognition performance than acuity. Acuity was more variable than fixation stability in patients with (at least partial) foveal vision (Fig. 6). Subjects with foveal vision all showed stable fixation (100% within 2°); however, their visual acuity ranged from 20/20 to 20/70. Fixation stability is important for

visual performance in a range of tasks; for example, unstable fixation is associated with slower reading speeds.⁵⁵ Patients with retinitis pigmentosa (in which the macula is spared) and patients with Stargardt dystrophy (in which the macula is diseased) performed equivalently poorly in a famous face recognition task, despite showing markedly different visual acuity.²² Potentially, fixation stability may be a more useful indicator of face recognition problems than acuity, despite that both are associated with performance in our study.

Recent work^{15,56–58} suggests that more general eye movement behavior (not just fixation stability) is crucial for visual tasks in patients with visual field loss. For example, patients with AMD with an established PRL show a different fixation pattern than controls when looking at a face; patients placed fewer fixations on the internal features (e.g., eyes) of a target face than controls.⁵⁶ Eye movement control training can improve performance on visual tasks such as reading⁵⁷ and potentially also improve face perception. However, in many real-world situations, people would have the opportunity to view a face for longer than in our study, so fixation stability may be less important for face perception tasks when faces are presented for extended durations. Conversely, fast face recognition may be required to comprehend television or films, particularly those with multiple and rapid scene cuts.

The main finding of our study is summarized in Figure 4: the LV group performed worse than the other groups at all blur levels. What does our modeling add to this? First, it allows us to quantify the size of the differences between groups, and the performance changes as a function of blur and block. Second, analyzing the data in a multilevel framework means that our inferences for groups with fewer participants are likely to be more robust because these estimates will be pulled toward the population mean (“shrinkage”) to the extent that the data suggest. Third and most important, estimating the posterior probabilities over the parameter space using Bayesian inference has a number of desirable properties over traditional null hypothesis significance testing (NHST). For the present application, three pertinent advantages of the Bayesian approach over NHST are that inferences over the posterior do not require corrections for making multiple comparisons, inferences do not depend on the experimenters’ intentions for stopping data collection, and Bayesian analyses coherently handle unbalanced designs by propagating uncertainty through all levels of the model. Interested readers are referred to other studies^{36,49,59–65} and to textbooks^{37,48} for recent introductions to and discussions of this type of inference, which is gaining traction in many scientific fields.

As many in the field of vision rehabilitation have stated (e.g., Trauzettel-Klosinski⁶⁶), practitioners must understand the impact of each patient’s vision loss on various visual functions to offer appropriate vision rehabilitation. Our patient sample included patients with very different pathologies and visual function (Table 1). While a variety of pathologies impair face recognition (see Dulin et al.²²), our results suggest that the key determinant of face recognition performance was whether the fovea was preserved or impaired. Patients in the LV:F group recognized as many faces as controls, despite significant field loss. While our experiment measured face recognition performance in an objective paradigm with high sensitivity (10-alternative forced-choice task), the strength of the conclusions we can draw is limited by our small sample of patients. An independent replication of our result in a larger sample of patients would strengthen the conclusion of this study that patients with preserved foveal vision (but with other visual deficits) can recognize faces to the same extent as someone with healthy vision.

CONCLUSIONS

We examined the effects of image blur on face identification in patients with low vision. Patients with at least partially preserved foveal vision and stable fixation performed similarly to normally sighted control observers, and face identification steadily decreased as stimulus blur increased. Patients with no foveal vision and unstable, eccentric fixation were severely impaired in the task at all levels of image blur. The results suggest that face recognition performance is limited by the highest spatial frequencies that are available to the observer, and this may be predicted by visual acuity and fixation stability as assessed in the clinic.

Acknowledgments

The authors thank William H. Seiple for comments and suggestions and Joy Facella-Ervolini for assistance with data collection.

TSAW was supported by Fellowship 634560 from the National Health and Medical Research Council of Australia. PJB was supported by Grants EY019281 and EY018664 from the National Institutes of Health.

Disclosure: **T.S.A. Wallis**, None; **C.P. Taylor**, None; **J. Wallis**, None; **M.L. Jackson**, None; **P.J. Bex**, Adaptive Sensory Technology, LLC (S), P

References

1. Sekuler AB, Gaspar CM, Gold JM, Bennett PJ. Inversion leads to quantitative, not qualitative, changes in face processing. *Curr Biol*. 2004;14:391–396.
2. Owsley C, McGwin G, Lee PP, Wasserman N, Searcey K. Characteristics of low-vision rehabilitation services in the United States. *Arch Ophthalmol*. 2009;127:681–689.
3. Watson GR, De l’Aune W, Stelmack J, Maino J, Long S. National survey of the impact of low vision device use among veterans. *Optom Vis Sci*. 1997;74:249–259.
4. Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. Human photoreceptor topography. *J Comp Neurol*. 2004;292:497–523.
5. Curcio CA, Allen KA. Topography of ganglion cells in human retina. *J Comp Neurol*. 1990;300:5–25.
6. Robson JG, Graham N. Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Res*. 1981;21:409–418.
7. Rovamo J, Virsu V, Näsänen R. Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature*. 1978;271:54–56.
8. Hilz R, Cavanaugh CR. Functional organization of the peripheral retina: sensitivity to periodic stimuli. *Vision Res*. 1974;14:1333–1337.
9. Kelly DH. Retinal inhomogeneity. I: spatiotemporal contrast sensitivity. *J Opt Soc Am A*. 1984;1:107–113.
10. Johnston A. Spatial scaling of central and peripheral contrast-sensitivity functions. *J Opt Soc Am A*. 1987;4:1583–1593.
11. Kwon M, Legge GE. Spatial-frequency cutoff requirements for pattern recognition in central and peripheral vision. *Vision Res*. 2011;51:1995–2007.
12. Bullimore MA, Bailey IL, Wacker RT. Face recognition in age-related maculopathy. *Invest Ophthalmol Vis Sci*. 1991;32:2020–2029.
13. Barnes CS, De l’Aune W, Schuchard RA. A test of face discrimination ability in aging and vision loss. *Optom Vis Sci*. 2011;88:188–199.
14. Glen FC, Crabb DP, Smith ND, Burton R, Garway-Heath DE. Do patients with glaucoma have difficulty recognizing faces? *Invest Ophthalmol Vis Sci*. 2012;53:3629–3637.

15. Glen FC, Smith ND, Crabb DP. Saccadic eye movements and face recognition performance in patients with central glaucomatous visual field defects. *Vision Res.* 2013;82:42–51.
16. Peli E, Goldstein RB, Young GM, Trempe CL, Buzney SM. Image enhancement for the visually impaired: simulations and experimental results. *Invest Ophthalmol Vis Sci.* 1991;32:2337–2350.
17. Tejeria L, Harper RA, Artes PH, Dickinson CM. Face recognition in age related macular degeneration: perceived disability, measured disability, and performance with a bioptic device. *Br J Ophthalmol.* 2002;86:1019–1026.
18. Yu D, Chung STL. Critical orientation for face identification in central vision loss. *Optom Vis Sci.* 2011;88:724–732.
19. Mei M, Leat SJ. Quantitative assessment of perceived visibility enhancement with image processing for single face images: a preliminary study. *Invest Ophthalmol Vis Sci.* 2009;50:4502–4508.
20. Boucart M, Dinon JF, Despretz P, Desmettre T, Hladiuk K, Oliva A. Recognition of facial emotion in low vision: a flexible usage of facial features. *Vis Neurosci.* 2008;25:603–609.
21. Peli E, Lee E, Trempe CL, Buzney S. Image enhancement for the visually impaired: the effects of enhancement on face recognition. *J Opt Soc Am A.* 1994;11:1929–1939.
22. Dulin D, Cavezian C, Serrière C, Bachoud-Levi AC, Bartolomeo P, Chokron S. Colour, face, and visuospatial imagery abilities in low-vision individuals with visual field deficits. *Q J Exp Psychol (Hove).* 2011;64:1955–1970.
23. Gaspar C, Sekuler AB, Bennett PJ. Spatial frequency tuning of upright and inverted face identification. *Vision Res.* 2008;48:2817–2826.
24. Hussain Z, Sekuler AB, Bennett PJ. Perceptual learning modifies inversion effects for faces and textures. *Vision Res.* 2009;49:2273–2284.
25. Hussain Z, Sekuler AB, Bennett PJ. Contrast-reversal abolishes perceptual learning. *J Vis.* 2009;9:20.1–20.8.
26. Pachai MV, Sekuler AB, Bennett PJ. Practice with inverted faces selectively increases the use of horizontal information. *J Vis.* 2012;12:975–975.
27. Goffaux V, Dakin SC. Horizontal information drives the specific signatures of face processing [serial online]. *Front Psychol.* 2010;1:143. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3153761/>. Accessed December 7, 2013.
28. Dakin SC, Watt RJ. Biological “bar codes” in human faces. *J Vis.* 2009;9:2.1–2.10.
29. Schuchard RA. Preferred retinal loci and macular scotoma characteristics in patients with age-related macular degeneration. *Can J Ophthalmol.* 2005;40:303–312.
30. Jackson M, Crossland MD, Wallis J, Seiple W. Microperimetry in rehabilitation. In: Mastropasqua L, Vingolo EM, Midena E, Jackson M, Mark IS, eds. *Handbook of Microperimetry in Visual Rehabilitation*. Padova, Italy: Nidek Technologies; 2013.
31. Kleiner M, Brainard D, Pelli DG. What’s new in Psychtoolbox-3? *Perception.* 2007;36:ECVP Abstract Supplement.
32. Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vis.* 1997;10:437–442.
33. Tyler CW. Colour bit-stealing to enhance the luminance resolution of digital displays on a single pixel basis. *Spatial Vis.* 1997;10:369–377.
34. Gold JM, Bennett PJ, Sekuler AB. Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Res.* 1999;39:3537–3560.
35. Crossland MD, Dunbar HMP, Rubin GS. Fixation stability measurement using the MP1 microperimeter. *Retina.* 2009;29:651–656.
36. Gelman A. Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics.* 2006;48:432–435.
37. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press; 2007.
38. Moscatelli A, Mezzetti M, Lacquaniti F. Modeling psychophysical data at the population-level: the generalized linear mixed model. *J Vis.* 2012;12:26, 1–17.
39. Knoblach K, Maloney LT. *Modeling Psychophysical Data in R*. New York: Springer; 2012.
40. Cheung SH, Kallie CS, Legge GE, Cheong AMY. Nonlinear mixed-effects modeling of MNREAD data. *Invest Ophthalmol Vis Sci.* 2008;49:828–835.
41. Stan Development Team. *Stan: A C++ Library for Probability and Sampling* [computer program]. Version 2.0.1. 2013. Available at: <http://mc-stan.org/>. Accessed December 8, 2013.
42. Hoffman MD, Gelman A. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Machine Learning Res.* In press.
43. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing; 2011.
44. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2009.
45. Wickham H. The split-apply-combine strategy for data analysis. *J Stat Software.* 2011;40:1–29.
46. Xie Y. knitr: A comprehensive tool for reproducible research in R. In: Stodden V, Leisch F, Peng RD, eds. *Implementing Reproducible Computational Research*. Boca Raton, FL: Chapman & Hall/CRC; 2013.
47. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci.* 1992;7:457–472.
48. Kruschke JK. *Doing Bayesian Data Analysis: A Tutorial With R and BUGS*. San Diego: Academic Press/Elsevier; 2011.
49. Kruschke JK. Bayesian estimation supersedes the *t* test. *J Exp Psychol Gen.* 2013;142:573–603.
50. Bouma H. Interaction effects in parafoveal letter recognition. *Nature.* 1970;226:177–178.
51. Galvin SJ, O’Shea RP, Squire AM, Govan DG. Sharpness overconstancy in peripheral vision. *Vision Res.* 1997;37:2035–2039.
52. von Noorden GK, Mackensen G. Phenomenology of eccentric fixation. *Am J Ophthalmol.* 1962;53:642–660.
53. Timberlake GT, Mainster MA, Peli E, Augliere RA, Essock EA, Arend LE. Reading with a macular scotoma, I: retinal location of scotoma and fixation area. *Invest Ophthalmol Vis Sci.* 1986;27:1137–1147.
54. White JM, Bedell HE. The oculomotor reference in humans with bilateral macular disease. *Invest Ophthalmol Vis Sci.* 1990;31:1149–1161.
55. Crossland MD, Culham LE, Rubin GS. Fixation stability and reading speed in patients with newly developed macular disease. *Ophthalmic Physiol Opt.* 2004;24:327–333.
56. Seiple W, Rosen RB, Garcia PMT. Abnormal fixation in individuals with age-related macular degeneration when viewing an image of a face. *Optom Vis Sci.* 2013;90:45–56.
57. Seiple W, Grant P, Szlyk JP. Reading rehabilitation of individuals with AMD: relative effectiveness of training approaches. *Invest Ophthalmol Vis Sci.* 2011;52:2938–2944.
58. Seiple W. Eye-movement training for reading in patients with age-related macular degeneration. *Invest Ophthalmol Vis Sci.* 2005;46:2886–2896.
59. Gelman A, Hill J, Yajima M. Why we (usually) don’t have to worry about multiple comparisons. *J Res Educ Effectiveness.* 2012;5:189–211.

60. Kruschke JK. Bayesian data analysis. *WIREs Cogni Sci.* 2010;1: 658–676.
61. Kruschke JK. What to believe: Bayesian methods for data analysis. *Trends Cogn Sci.* 2010;14:293–300.
62. Kruschke JK. Bayesian assessment of null values via parameter estimation and model comparison. *Perspect Psychol Sci.* 2011; 6:299–312.
63. Wagenmakers EJ. A practical solution to the pervasive problems of *P* values. *Psychon Bull Rev.* 2007;14:779–804.
64. Wetzels R, Raaijmakers JGW, Jakab E, Wagenmakers EJ. How to quantify support for and against the null hypothesis: a flexible WinBUGS implementation of a default Bayesian *t* test. *Psychonomic Bull Rev.* 2009;16:752–760.
65. Wetzels R, Matzke D, Lee MD, Rouder JN, Iverson GJ, Wagenmakers EJ. Statistical evidence in experimental psychology: an empirical comparison using 855 *t* tests. *Perspect Psychol Sci.* 2011;6:291–298.
66. Trauzettel-Klosinski S. Rehabilitation for visual disorders. *J Neuroophthalmol.* 2010;30:73–84.