Computational-observer analysis of illumination discrimination

Xiaomao Ding

Neuroscience Graduate Group,
University of Pennsylvania, Philadelphia, PA, USA ✉

Ana Radonjić

Department of Psychology, University of Pennsylvania,
Philadelphia, PA, USA ✉

Nicolas P. Cottaris

Department of Psychology, University of Pennsylvania,
Philadelphia, PA, USA ✉

Haomiao Jiang

Department of Electrical Engineering,
Stanford University, Sunnyvale, CA, USA
Current address: Google Research ✉

Brian A. Wandell

Department of Psychology, Stanford University,
Sunnyvale, CA, USA ✉

David H. Brainard

Department of Psychology, University of Pennsylvania,
Philadelphia, PA, USA ✉

The spectral properties of the ambient illumination provide useful information about time of day and weather. We study the perceptual representation of illumination by analyzing measurements of how well people discriminate between illuminations across scene configurations. More specifically, we compare human performance to a computational-observer analysis that evaluates the information available in the isomerizations of cone photopigment in a model human photoreceptor mosaic. The performance of such an observer is limited by the Poisson variability of the number of isomerizations in each cone. The overall level of Poisson-limited computational-observer sensitivity exceeded that of human observers. This was modeled by increasing the amount of noise in the number of isomerizations of each cone. The additional noise brought the overall level of performance of the computational observer into the same range as that of human observers, allowing us to compare the pattern of sensitivity across stimulus manipulations. Key patterns of human performance were not accounted for by the computational observer. In particular, neither the elevation of illumination-discrimination thresholds for illuminant changes in a blue color direction (when thresholds are expressed in CIELUV ΔE units), nor the effects of varying the ensemble of surfaces in the scenes being viewed, could be accounted for by variation in the information available in the cone isomerizations.

Introduction

The spectral properties of the ambient illumination provide useful information about time of day and weather. Indeed, variation in natural illumination spectra occurs during the course of the day (e.g., Hernandez-Andres, Romero, Nieves, & Lee, 2001; Spitschan, Aguirre, Brainard, & Sweeney, 2016), within single natural scenes (Nascimento, Amano, & Foster, 2016), and when we move between natural and artificial illumination (Wyszecki & Stiles, 1982). A number of psychophysical paradigms have been developed to study the perceptual representation of illumination (Kardos, 1928; Beck, 1959; Oyama, 1968; Kozaki & Noguchi, 1976; Gilchrist & Jacobsen, 1984; Noguchi & Kozaki, 1985; Hurlbert, 1989; Logvinenko & Menshikova, 1994; Rutherford & Brainard, 2002; Logvinenko

& Maloney, 2006; Lee & Brainard, 2011). Recently, a number of labs have reported psychophysical threshold measurements of how well people can discriminate between illuminations across scene configurations (Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014; Radonjić et al., 2016; Weiss & Gegenfurtner, 2016; Alvaro, Linhares, Moreira, Lillo, & Nascimento, 2017; Radonjić et al., 2018; Aston, Radonjić, Brainard, & Hurlbert, 2019; see also Lucassen, Gevers, Gijsenij, & Dekker, 2013).

A key step in interpreting psychophysical threshold measurements is to understand the degree to which patterns in the data are driven by variation in the information available from the stimuli. This has often been accomplished by comparing human performance to that of an *ideal observer* that makes optimal use of the task-relevant information available in the stimulus or at some early stage of the visual processing (Barlow, 1962; Green & Swets, 1966; Banks, Geisler, & Bennett, 1987; Geisler, 1989; Sekiguchi, Williams, & Brainard, 1993; Geisler, 2011; Cottaris, Jiang, Ding, Wandell, & Brainard, 2018). Such ideal observer analyses clarify which aspects of performance may be accounted for by the properties of the stimuli and well-understood mechanisms of early vision.

In this paper, we review previously reported measurements of human psychophysical performance on an illumination-discrimination task (Radonjić et al., 2016; Radonjić et al., 2018). We then compare human performance to that of a computational observer that uses the information available in the cone photopigment isomerizations to perform the same task. We ask whether the stimulus dependent changes in human performance across different illumination changes and scene configurations are consequences of differences in the information available to perform the task. The analysis clarifies which aspects of performance require additional explanation in terms of the action of visual mechanisms beyond the isomerization of photopigment by light.

Our approach shares much with the ideal observer analysis, but rather than using an analytic calculation to estimate ideal performance levels, we employ computer simulations and machine learning. For this reason, we refer to our approach as a *computational-observer analysis* (Farrell, Jiang, Winawer, Brainard, & Wandell, 2014; Jiang et al., 2017; Cottaris et al., 2018; cf. Lopez, Murray, & Goodenough, 1992). Unlike the ideal observer, where the decision rule is implemented on the assumption that the observer has perfect information about the statistical properties of the stimuli, the computational observer must learn these properties from a large number of training samples.

Illumination-discrimination psychophysics: Modeled Experiment 1

We analyze two similar illumination-discrimination experiments; we begin by providing an overview of the first. Detailed methods and results for this experiment (Modeled Experiment 1) are reported elsewhere (Radonjić et al., 2018, experiment 2 in that paper, fixed-surfaces condition), so our overview is brief. The second experiment (Modeled Experiment 2) is similar in design and is also reported in detail elsewhere (Radonjić et al., 2016).

Both modeled experiments measured humans' ability to discriminate changes in illumination across four illumination-change directions—blue and yellow (which were aligned with the daylight locus) and red and green (which were orthogonal to the daylight locus). On each trial of the experiments, observers viewed three successive computer-generated images, displayed on a calibrated color monitor. The scene geometry was held fixed across the experiment, and all the surfaces in the scene remained unchanged: Only the spectral power distribution of the illumination varied. The first image on a trial was presented in the *reference interval* (2,370 ms). For this image, the scene was illuminated by the *target illumination*, a metamer[1] for natural daylight with a color temperature of approximately 6700 K. The reference interval was followed by two *comparison intervals*. The comparison intervals (870 ms) were separated from each other and from the reference interval by interstimulus intervals (750 ms). During each comparison interval, an image of the scene was again shown. In one comparison interval, the scene was illuminated by the target illumination and in the other by a *test illumination*. The order of the two comparisons was randomized on each trial. The observer's task was to choose which of the two comparison intervals had scene illumination most similar to the target illumination. On each trial, the test illumination was chosen from a pool of 200 prespecified illuminations. These varied in steps of approximately 1 CIELUV ΔE unit (CIE, 2004) in each of four illumination-change directions: blue, yellow, green, and red (50 steps per direction).[2] Figure 1 shows example images of the scene under the target illumination and four test illuminations.

During each session, trials probing the four illumination-change directions were interleaved. The illumination-change steps were determined through twelve interleaved 1-up-2-down staircases (three independent staircases for each illumination-change direction). The illumination step size for a staircase was decreased when the observer responded correctly twice in a row on trials governed by that staircase, and increased each
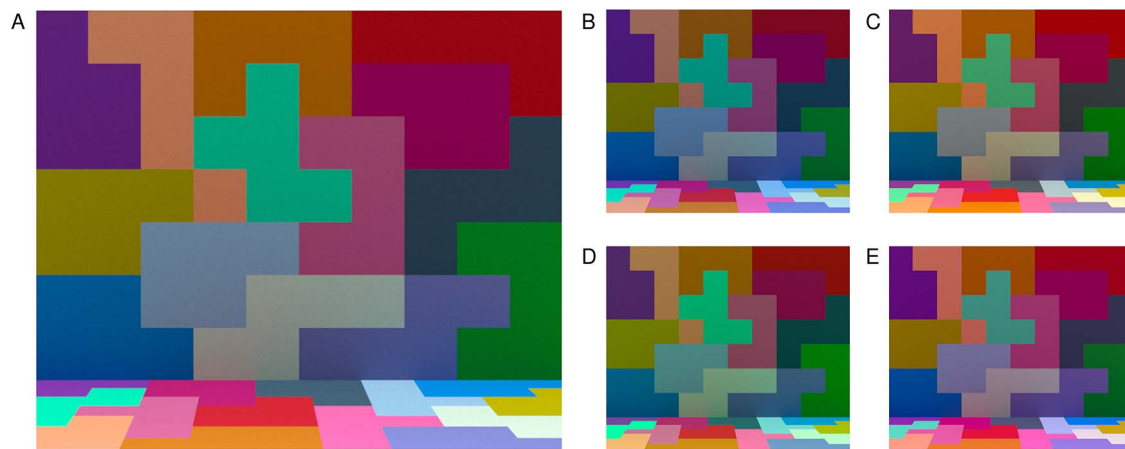
Figure 1. Experimental stimuli. (A) Image of the scene illuminated by the target illumination. (B–E) Images of the scene illuminated by four test illuminations. Each illumination is approximately 30 ΔE steps away from the target. Panels (B–E) correspond to test illuminations in the blue, yellow, green, and red illumination-change directions respectively. Images in (B–E) are shown at a smaller size than in (A) to conserve space. In the experiment, the target and test illumination images were the same size. Images rendered here and in other figures in this paper are illustrative. The rendering process is unlikely to preserve the precise color appearance of the experimental stimuli.

time the observer responded incorrectly. Each staircase terminated after the eighth reversal.

Observers' thresholds to discriminate illumination changes were measured for each of the four directions, relative to the target illumination. Thresholds corresponded to 70.71% accuracy and were determined using a maximum likelihood fit to the combined data for all three staircases in each illumination-change direction. The 70.71% correct criterion was used because this is the performance convergence point for the staircases (Wetherill & Levitt, 1965). Observers completed two experimental sessions, and thresholds obtain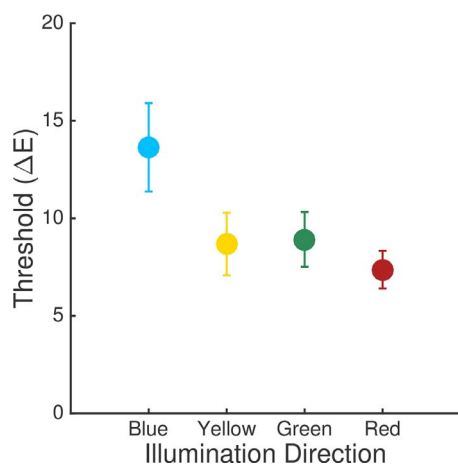ed in each session were averaged for each observer. Each observer's right eye was tracked using an Eyelink 1000 (SR Research, Ltd.), which allowed us to record eye fixation positions on each trial.

Figure 2 illustrates the average thresholds across ten observers. The average threshold for the blue direction is elevated relative to those obtained for the other illumination-change directions, when thresholds are plotted using the CIELUV ΔE metric. The elevation for thresholds in the blue direction has been found in other studies with a similar design, for scenes where the average reflectance of the surfaces in the scene is close to that of a neutral nonselective gray (Pearce et al., 2014; Radonjić et al., 2016). Two essentially identical experiments reported with this one (Radonjić et al., 2018, experiments 1 and A1 in that paper, fixed-surfaces condition) yielded similar results. We are particularly interested here in whether the elevation of threshold in the blue direction has its roots in a relative paucity of information about this illumination change in the representation of the visual scene provided by the cone photoreceptor mosaic, which could arise because of the relatively small number of S cones in the mosaic.

Figure 2. Experimental results. Thresholds averaged across observers for Experiment 1. Error bars are ±1 *SEM*. Each point shows the threshold for one illumination-change direction. Data originally reported in Radonjić et al. (2018, experiment 2 in that paper, fixed-surfaces condition).

Computational-observer analysis

We analyze the information available for performing the illumination-discrimination task present in the photopigment isomerizations of the cone photoreceptors. We used the Image Systems Engineering Toolbox for Biology (ISETBio; isetbio.org) to model retinal image formation and the absorption of light in the foveal cone mosaic. ISETBio provides functions that model early vision and allow computation of the visual
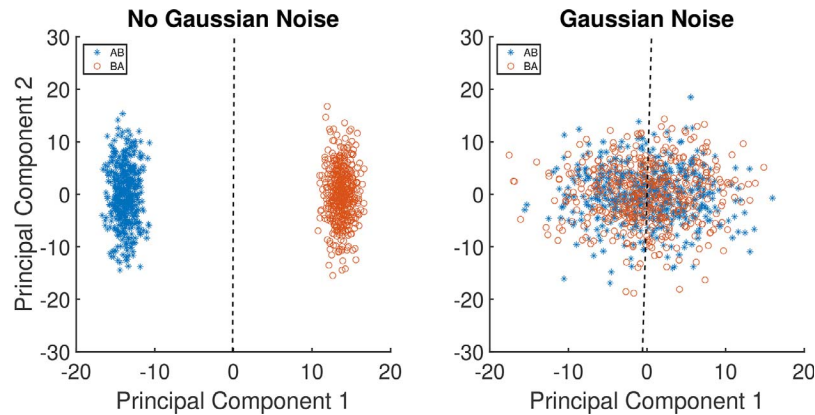
Figure 3. Examples of learned classification boundaries. (A) Plot of 100 samples of AB data vectors and 100 samples BA data vectors projected onto the first two principal components of the cone mosaic response space. Data are for the illumination-change step size $\Delta E = 1$ in the blue direction. The dashed line is the linear SVM decision boundary learned from the 200 samples in this training set. The variability is due to Poisson noise; no Gaussian noise was added. (B) Corresponding plot with 15× additive Gaussian noise added to the responses, prior to data reduction by PCA and classification learning. Note that the principal components are not identical between the two panels, as the PCA calculation was done separately on the two training sets.

system's responses to calibrated stimuli at various stages of processing. It includes routines that compute the photopigment isomerizations of each cone in a patch of photoreceptor mosaic, accounting for the scene radiance, lens and macular pigment transmissivities, image formation by the eye's optics, and the spectral sensitivities and interleaved sampling of the three classes of cones.

ISETBio is written in MATLAB (MathWorks, Natick, MA) and is freely available under an open-source software license (isetbio.org). The code and data required to use ISETBio to reproduce the calculations and figures in this paper are available at https://github.com/isetbio/BLIlluminationDiscriminationCalcs. This repository also includes routines that interface between the ISETBio representations and MATLAB's support vector machine (SVM) implementation, which we use to implement our computational observer.

Modeling of the stimuli

The stimuli were presented on calibrated computer monitors with known size and distance from the observer. In Experiment 1, stimuli were presented to the right eye only and we modeled the information contained in the right eye image. In Experiment 2, the stimuli were presented stereoscopically, but for simplicity we modeled only the information contained in the left eye image. We imported the RGB stimulus images into the ISETBio *scene* format, which uses the display calibration information to compute the spectral power distribution of the light emitted from each location in the image. The scene representation also specifies the size of the image and the distance between

the display and the eye. These values were set to match their experimental values.[3]

The retinal image

We used ISETBio routines to compute the retinal image from each stimulus image. In ISETBio terminology, this is referred to as the *optical image*. These routines incorporate the size of the pupil (6 mm in our calculations, with this size being estimated from the luminance of our stimuli using the formulae in Watson & Yellott, 2012), the geometry of image formation, absorption of light by the lens, and blurring by the eye's optics. We used the ISETBio default estimates of lens density from Bone, Landrum, and Cairns (1992) and the polychromatic shift-invariant wavefront aberration-derived point spread function (PSF) of a typical subject from the Thibos, Hong, Bradley, and Cheng (2002) dataset. Cottaris et al. (2018) describe the method used to select the typical subject from the larger dataset.

Cone photopigment isomerizations

We estimated the mean number of photopigment isomerizations in the cones for a $1.1° \times 1.1°$ patch of foveal retina. We chose small patches of retina because restricting the analysis to small patches makes calculation of computational-observer performance tractable given the current implementation and computing power. The full analysis is thus achieved by breaking the image down into a set of patches and aggregating performance over these patches. Figure 4A below
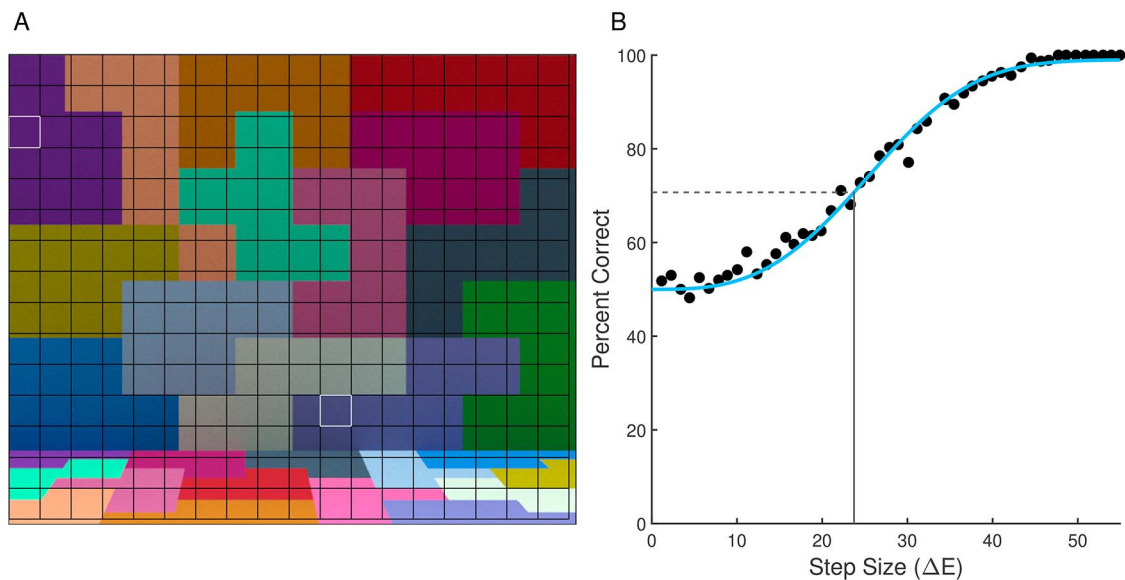
A

B

Figure 4. Obtaining computational-observer thresholds. (A) Stimulus for Modeled Experiment 1 with uniform grid superimposed on top. Each square patch in the grid is $1.1° \times 1.1°$ visual angle. Each patch was analyzed separately. (B) Plot of linear SVM performance at each stimulus level for one stimulus patch (row three, column one); white outline in (A). Points represent the performance of a linear SVM trained and tested for each step size for the blue direction with 20× Gaussian noise factor. The best-fit Weibull psychometric function is depicted with the solid blue line. The horizontal dashed gray line is the 70.71% accuracy level, and the vertical solid gray line is the threshold (23.7 ΔE units in this example), calculated using the inverse of the fit function.

illustrates the layout of the patches relative to the stimuli for Modeled Experiment 1.

The model cone mosaic was a hexagonal mosaic in which cone density, cone aperture, and outer segment length all varied with eccentricity according to measurements by Curcio et al. (1991). The construction of such mosaics is described in detail in Cottaris et al. (2018). The peak cone density was 204,213 cones/mm², with a corresponding minimum cone spacing of 2.36 microns, which is within the range of foveal densities reported by Curcio et al. (1991). There were 9629 cones in all, 6167 (64%) L cones, 3,085 (32%) M cones, and 377 (4%) S cones, corresponding to current views on the relative numbers of cones of each class in the typical human retina (Hofer, Carroll, Neitz, Neitz, & Williams, 2005). There were no S cones in the central 0.15°, with a semiregular arrangement of S cones outside this region. The non-S cones were assigned as L or M randomly with 2:1 as the expected L:M ratio. Photopigment absorbance for the L, M, and S cones was taken from the CIE 2006 standard (CIE, 2006), and absorption by the foveal macular pigment (Stockman, Sharpe, & Fach, 1999) was incorporated. These choices, together with that used for lens density above, yield the CIE (2006) 2° cone fundamentals. We assumed a quantal efficiency for cone photopigment of 0.67 (Rodieck, 1998) and a foveal cone inner segment acceptance area of 1.96 um², an area that sits between those provided by Kolb (http://webvision.med.utah.edu) and Rodieck (1998). We assumed a mean rate of spontaneous

isomerizations of 100/cone-sec, consistent with estimates obtained psychophysically but lower than those obtained physiologically (see Koenig & Hofer, 2011). We used a cone integration time of 50 ms and computed the mean number of isomerizations of each cone in response to each image. From the mean number of isomerizations, we can simulate the number of isomerizations on an individual trial, because trial-by-trial isomerizations are Poisson distributed (Hecht, Schlaer, & Pirenne, 1942; Rodieck, 1998). The mean numbers of L, M, and S cone isomerizations under the standard illuminant for this experiment, taken across the image patches, were 439.8, 315.0, and 50.8 respectively for Modeled Experiment 1. For Modeled Experiment 2, the values were 325.6, 246.3, and 33.4. Supplementary Table S1 provides the mean number of isomerizations for all of the illuminations for both modeled experiments.

Obtaining computational-observer thresholds

Given the responses of a simulated cone mosaic to the experimental stimuli, we made computational-observer predictions of discrimination performance for each illumination-change direction and step size. To do so, we created a training set consisting of simulated trials and used this to learn a linear classifier that separated trials according to which comparison interval corresponded to the target. Each instance in the

training set had independently drawn noise added to the mean number of isomerizations for each cone. We predicted performance by evaluating the classifier on a separate test set (cross-validation). The test set consisted of additional simulated trials, each with independently drawn noise.

We studied performance for individual $1.1° \times 1.1°$ patches of foveal mosaic with a single fixation location at the center of each patch, which was held constant across the two comparison intervals. This analysis quantifies the information available from single patches of the stimulus. We then investigated how the information varies across stimulus patches as well as aggregated across patches. Our approach ignores the additional information provided in the reference interval. The motivation for including that interval in the experiment was to reduce observer's uncertainty about which of the two comparison-interval images corresponds to the target. Our computational observer was designed so that such uncertainty is not an issue.

For each illumination-change direction and step size, we calculated the mean number of isomerizations in the cones of the mosaic to the target-illumination image patch and the test-illumination image patch. To each mean response, we added Poisson noise to simulate trial-to-trial variability in the number of isomerizations in a single integration time. Then, we concatenated the resulting two response vectors into a single vector, with the target coming either first or second. We refer to simulated trials where the target vector comes first and the comparison vector second as AB trials, and simulated trials where the comparison vector comes first and the target second as BA trials. The task of the computational observer was to classify AB versus BA response vectors.

The training and test sets each consisted of 1,000 labeled concatenated response vectors of this sort (500 AB and 500 BA). We used the training set to learn a linear classifier, i.e., to find the hyperplane which best separated the two classes (AB versus BA) of concatenated response vectors. We did this using the support vector machine (SVM) algorithm (Manning, Raghavean, & Schutze, 2008), as implemented in MATLAB (function fitcsvm). This algorithm finds the hyperplane that maximizes the margin between the exemplars of the two classes, where margin refers to the amount of space without any data points around the classification boundary. Support vector machines provide an effective tradeoff between good classification performance on the training set and good generalization.

The dimensionality of the concatenated response vectors is 19,258 (for two concatenated $1.1° \times 1.1°$ patches of foveal retina). We reduced this dimensionality using principal components analysis (PCA). We first standardized each dimension (i.e., response of one cone in one interval) of the training set using its sample mean and standard deviation. Then, we ran principal components analysis on the training set and projected the standardized response vectors onto the first 400 principal components for training and testing.

Figure 3A shows the projection of 100 AB and 100 BA vectors from one stimulus patch, using nominal step size $\Delta E = 1$ in the blue illumination-change direction, onto the first two principal components obtained for this direction and step size. The dashed line shows the decision boundary of a linear SVM trained on these vectors. Using this decision boundary led to performance of 100% on the training set as well as on the test set, indicating that a computational observer could perform perfectly on the task even for this smallest illumination-change step size. Since human thresholds considerably exceed $\Delta E = 1$, there are additional sources of noise beyond the Poisson variation in the number of isomerizations and/or inefficiencies in the way that the visual system uses the information provided by the cone photopigment isomerizations. Indeed, ideal observer studies of human discrimination performance for simple stimuli often find that ideal observers outperform human observers (Geisler, 1989). Because our goal is to understand whether the pattern of psychophysical performance we measured is driven by differences in information at the photoreceptor mosaic, we modeled the inefficiency of the postisomerization visual system by adding independent zero-mean Gaussian noise to each cone's response, in addition to the Poisson noise. Doing so degrades the performance of the computational observer, but in a manner that does not seem likely to introduce systematic changes in relative computational-observer performance across illumination-change directions. We set the variance of the added noise for each cone's response (both for the target and comparison responses) to a multiple of the mean response of the cones in the mosaic; we refer to the multiple chosen as the *noise factor*. The noise factor was varied systematically, providing us with a parameter that allowed us to match performance levels between computational and human observers. Expressing the magnitude of the added noise as a noise factor in the way we do relates the variance of the added noise to the mean variance of the Poisson noise intrinsic to photopigment isomerizations, and thus allows us to understand how much noise had to be added relative to that intrinsic Poisson noise.

Figure 3B shows the same 100 AB and BA vectors from 3A, but with 15× (noise factor) Gaussian noise added to the original responses. These noisy responses are projected onto principal components computed from a training set of response vectors with the same noise factor. The two classes are now overlapping, illustrating how adding the Gaussian noise reduces

performance. Indeed performance on the test case for this set is essentially at chance (51%).

To model the psychophysical data, we trained and tested SVMs for each combination of illumination direction, step size, and a series of noise factors. This provides us with a modeled percent correct for each combination. Noise factors were varied from 0 to 30 in steps of 5.[4]

As noted above, it was not computationally feasible for us to train the SVM's using a cone mosaic that captures the entire stimulus. For this reason, we partitioned the scene into $1.1° \times 1.1°$ square patches by imposing a uniform grid onto the stimulus, and then trained SVM's for each patch across all combinations of illumination direction, step size, and noise factor. Figure 4A illustrates the grid. On the right and bottom edges of the image, there are small regions of the stimulus not covered by the grid. These are neglected in the calculations. To obtain thresholds for each illumination-change direction and choice of noise factor for each $1.1° \times 1.1°$ patch, we analyze computational-observer performance as a function of step size by fitting a Weibull psychometric function. This relates computational-observer performance to illuminant-change step size. In performing the fit, the actual ΔE values for each illumination change, rather than the nominal values, were used (see Radonjić et al., 2018). The fit was obtained using the maximum-likelihood method implemented in the Palemedes Toolbox (Prins & Kingdom, 2018; www.palamedestoolbox.org, version 1.8.2). Figure 4B shows computational-observer performance and the fit, for the blue illumination-change direction at 20× noise factor, for the patch outlined in white in the upper left (third row from the top, first column from the left) of Figure 4A. Using the fitted psychometric functions, 70.71% correct thresholds were extracted.

We verified for one $1.1° \times 1.1°$ patch and one noise factor (4×), for a set of changed illuminations in the blue direction, that the PCA dimensionality reduction to 400 did not substantially distort the percent correct obtained by the SVM-based computational observer, compared to more time consuming calculations without dimensionality reduction. Supplementary Figure S1 shows percent correct plotted against illumination change ΔE, for the full 19,258 dimensional training/test vectors and for various choices of reduced dimensionality. The values for different dimensionalities are similar to each other and to the value obtained for the full dimensionality.

Figure 5A shows the computational-observer thresholds for each illumination direction plotted as a function of noise factor, for the same stimulus patch (row three, column one) whose performance for the blue illuminant-change direction is shown in Figure 4B. For low added noise, thresholds are very low, as one

would expect based on Figure 3A. As the noise factor increases, thresholds also increase. The relative ordering of the thresholds across illumination-change directions is preserved across noise factors (yellow > blue > green > red).

Figure 5B shows performance for a different stimulus patch, indicated by a white outline in the lower-center portion of Figure 4A (row 12, column 11). Thresholds again increase with noise factor, but here the ordering of thresholds in the different illumination-change directions is different than in 5A (green > red > yellow > blue). This effect is due to the difference in surface reflectance at the two patches. The two patches for which performance is illustrated in Figure 5A and B were chosen to illustrate the large effect that patch choice can have on computational-observer performance.

Because predicted performance depends on which patch is examined, to make overall predictions it is necessary to aggregate across the individual patches. There are a number of ways to implement such aggregation. Here, for each illumination direction, step size, and noise level, we averaged the percent correct performance obtained for each of the 270 patches (Figure 4A) to obtain a single aggregate psychometric function. We then analyzed this function using the same methods as for the individual patch data, to obtain an aggregate threshold as a function of noise factor. The results of this analysis are shown in Figure 5C. The aggregate performance more closely resembles the performance in 5A than in 5B, but differs in relative threshold order (blue > yellow > green > red). The aggregate thresholds represent an estimate of computational-observer performance when the information in all patches is weighted equally.

Relation to psychophysics

For each human observer, we found a single noise factor that minimized the sum of the squared error (across the four illumination-change directions) between the aggregate computational-observer thresholds (Figure 5C) and human thresholds. Linear interpolation was used to estimate computational-observer thresholds for noise factors between those for which performance was explicitly computed. The resulting computational-observer thresholds across observers were averaged to provide a fit to the average human data. This average fit is shown in Figure 6A. The computational-observer thresholds share with the human thresholds the elevation in the blue illumination-change direction relative to the red and green directions, indicating that the information available to the visual system at the retinal photoreceptor stage is already biased across illumination directions. However,
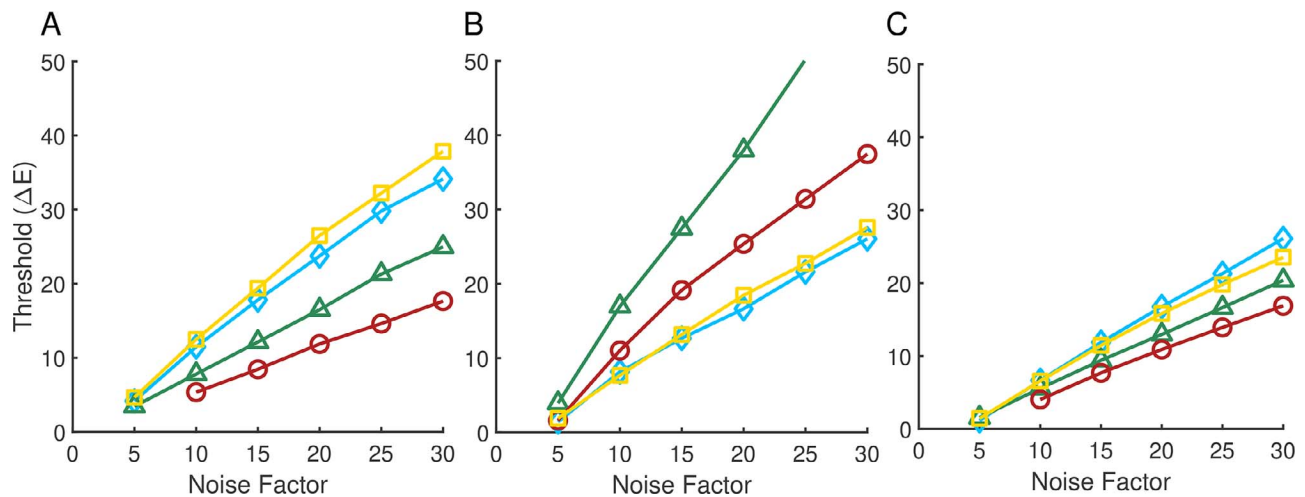
Figure 5. Computational-observer thresholds as a function of noise factor. (A) Computational-observer thresholds from a single stimulus patch (top left white outline in Figure 4A) for the four illumination-change directions (blue: blue diamonds; yellow: yellow squares; green: green triangles; red: red circles), as a function of noise factor. (B) Same as (A) but for a different stimulus patch (white outline in the lower center portion of Figure 4A). The line for the green illumination-change direction does not extend further because at higher noise levels performance was too poor to estimate a threshold. (C) Aggregated thresholds as a function of noise factor, obtained by averaging performance over all the patches shown in 4A for each illumination direction/step size/noise factor. The psychometric function used to obtain the aggregated thresholds was fit to the average performance values taken over all patches.

the difference in the computational-observer thresholds between the blue and yellow illumination-change directions is small and does not provide a clear explanation for the elevation of thresholds in the blue direction that is generally found in the human psychophysics, when thresholds are expressed using the CIELUV ΔE metric.

The average (across observers) noise factor obtained in the fitting above was 15.04. This factor provides an omnibus summary of the net decrease in performance between the computational and human observers. There are many factors that likely contribute to the decrease, including neural noise introduced after the cone photopigment isomerizations, differences in spatial and temporal integration of information between the computational and human observers, loss of information during the interstimulus intervals, and differences in the decision processes used by the computational and human observers. The present work
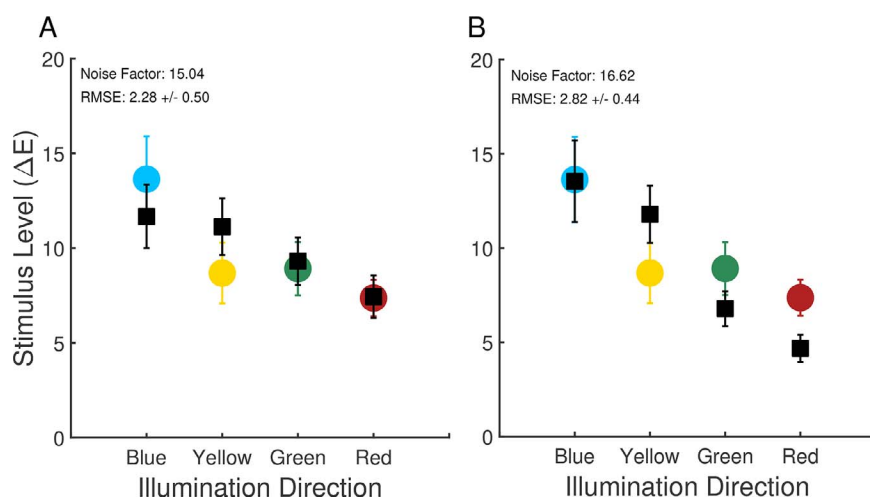

Figure 6. Comparison of computational and psychophysical performance: Experiment 1. (A) Computational-observer thresholds (black squares) along with human observer thresholds. The noise factor leading to the best fit to each human observer's data was found, and the resulting computational-observer thresholds were averaged over observers. Computational-observer thresholds were calculated using aggregation over all stimulus patches. Error bars are ± 1 *SEM* across observers. Data replotted from Figure 2. (B) Same as (A), except using computational-observer thresholds obtained with fixation-weighted aggregation.
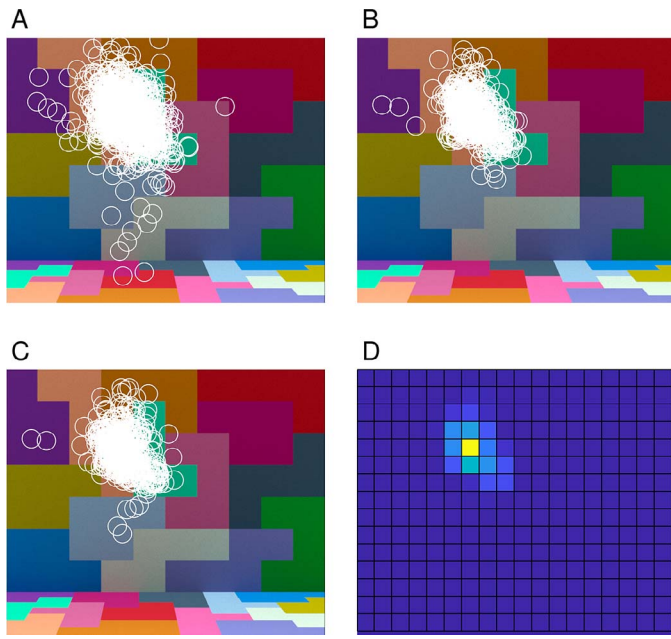
Figure 7. Eye fixation data. (A) Fixations made by one observer (observer eom in Radonjić et al., 2018; note that observer initials are fictional and do not provide identifying information) during the reference interval, aggregated over all trials for that observer in both sessions. (B) and (C) Fixations made by the same observer during the first and second comparison intervals. (D) Distribution of combined fixations from (B) and (C), represented as a heat map (yellow indicating a higher number of fixations and blue indicating a lower number). Each fixation was assigned to one of the $1° \times 1°$ stimulus patches (see Figure 4A).

does not distinguish the contribution of each of these factors.

Effect of eye fixations

The approach to aggregating the performance across patches described above weights the information from each stimulus patch equally. The fixations made by human observers during the experiment, however, show that they did not look at each part of the scene equally often. Figure 7A, 7B, and 7C illustrate the fixations made a by a single representative observer across both sessions of the experiment. Panel A shows the fixations made during the reference interval, while panels B and C show the fixations made during the first and second comparison intervals respectively. In all cases, the fixations generally cluster around one location in the scene, with the fixations made during the reference interval slightly more spread out, presumably because the duration of the reference interval was longer (2,370 ms vs. 870 ms). A more detailed analysis

of the fixation data is provided in Radonjić et al. (2018).

One approach to incorporating the fixation data into the computational-observer calculations is to weight each stimulus patch in the averaging step according to the fraction of fixations that landed within that patch during the two comparison intervals. Figure 7D illustrates the fixation-based weights for one observer as a heat map, with the weights for each patch obtained from the combined data shown in 7B and C. We incorporated fixation-based weights for each observer using that observer's eye fixation data and the computed thresholds for each illumination-change direction/step size/noise factor. We then found the noise factor for each observer that brought the computational-observer data into best register with that observer's data, and then averaged the thresholds across observers. The results are shown in Figure 6B. The effect of incorporating eye fixations is not large. The primary effect is the increase in the difference between the computational blue and yellow direction thresholds on the one hand and the red and green direction thresholds on the other. The difference between the computational observer blue and yellow direction thresholds also increased slightly, but re-mained considerably less than the difference in human thresholds for these directions. Once the noise factor was chosen to fit the psychophysical data, the RMSE of the fit was similar to but slightly worse (6B) than that obtained with the nonfixation-weighted analysis de-scribed above (6A).

Information provided by the different cone classes

The computational-observer thresholds are elevated in the blue and yellow directions relative to the red and green directions. This suggests that there is an asymmetry across illumination-change directions in the information available at the photoreceptor mosaic, when step size is expressed in ΔE units. We were curious about how the different cone classes contribute to this asymmetry. We recomputed computational-observer thresholds for six additional cone mosaics. These were mosaics with L cones only, M cones only, S cones only, L and S cones, L and M cones, and M and S cones. In constructing the dichromatic cone mosaics, missing L cones were replaced with M cones and vice-versa, whereas missing S cones were replaced with a mixture of L and M cones in a 2:1 (L:M) ratio.

Figure 8 shows computational-observer thresholds for the six additional mosaics (black circles in each plot) along with the corresponding computational-observer data for the original trichromatic mosaic (colored circles in each plot, replotted from black
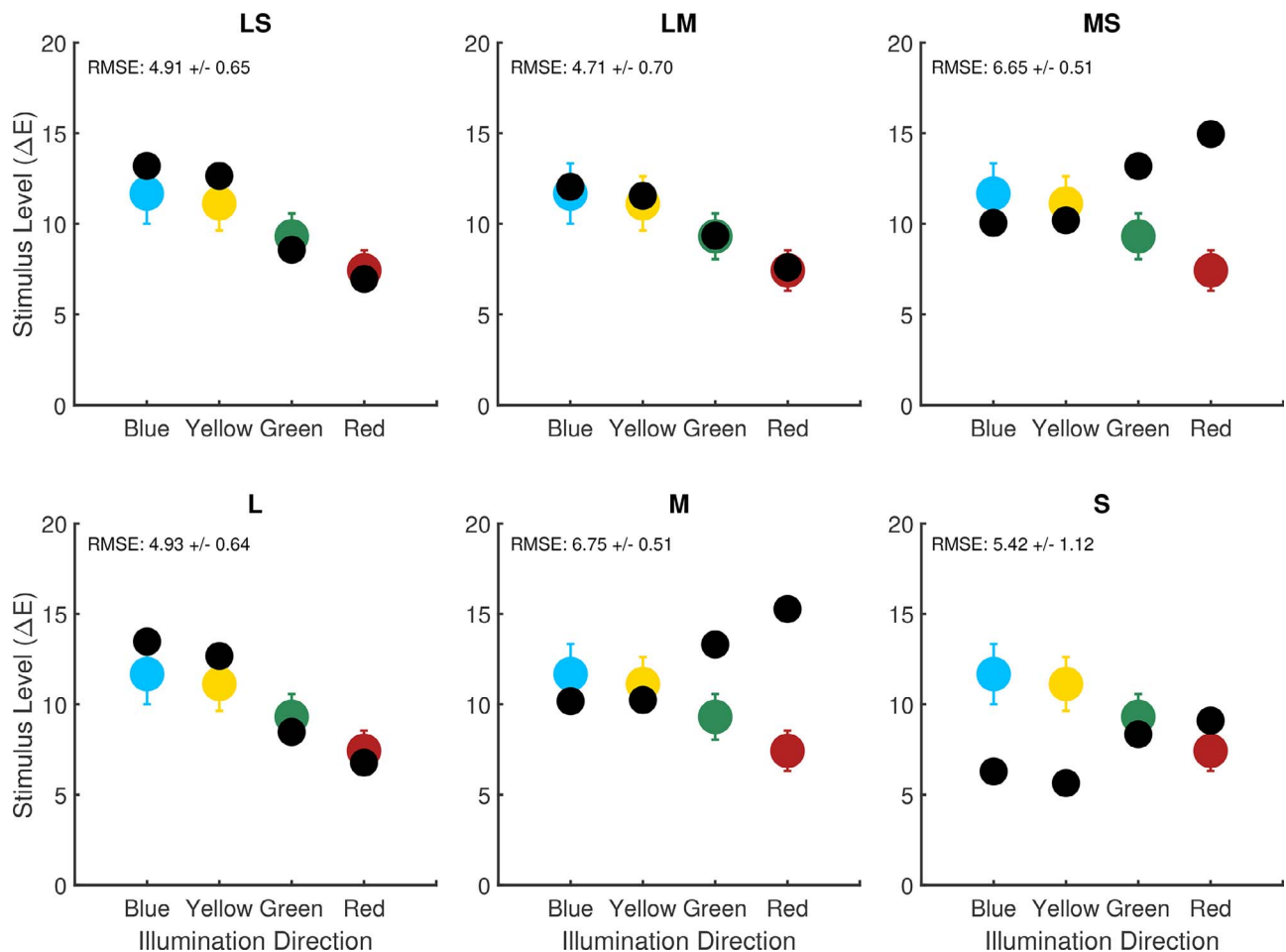
Figure 8. Effect of variation in mosaic. Best fit computational-observer thresholds for alternate mosaics (black circles in each panel) compared to the computational-observer thresholds for the trichromatic mosaic (colored circles in each panel, replotted from the black squares in Figure 6A). From left to right top row: LS cones only (no M, deuteranope), LM cones only (no S, tritanope), MS cones only (no L, protanope). From left to right bottom row: L cone only, M cone only, S cone only. Performance aggregation is over all stimulus patches, as we do not have fixation data corresponding to human observers with the alternate mosaics. The same noise factor (15.04) is used for all of the di- and monochromatic mosaic computational-observer thresholds shown here. This value is the average noise factor value obtained for the trichromatic computational-observer thresholds in 6A.

squares shown in Figure 6A). Computational-observer thresholds for the L, LS, and LM mosaics are similar to one another as well to those obtained with the original mosaic. This suggests that the elevation of computational-observer thresholds in the blue and yellow directions found with the original mosaic can be primarily attributed to the pattern of performance mediated by the L cones: mosaics containing L cones show the elevation, while those with only M and/or S cones do not. Possibly, the fact that L cones are the predominate type in the mosaic leads to the overall mosaic showing a pattern of performance that is most similar to the performance of the L-cone-only mosaic. This in turn might mean that some of the individual variability observed in illumination-discrimination performance (see Radonjić et al., 2016; Radonjić et al., 2018) could be related to individual variability in L:M cone ratio (Hofer et al., 2005; for a review see Brainard,

2015). We note, however, that there are other possibilities. For example, there is a systematic difference in the isomerization rate between L and S cones because of light absorption by the lens and macular pigment. This difference might interact with our choice of SVM linear classifier or our choice to model postisomerization processing with equal-variance Gaussian noise.

The relative computational-observer thresholds for M, S, and MS cone mosaics are also similar to one another, although sensitivity is lower overall for S, and these thresholds have a different pattern across illumination-change directions than that obtained with L-cone dominated mosaics. Here, computational-observer thresholds for the blue and yellow illumination directions are the lowest relative to the other color directions.

Aston, Turner, Le Couteur Bisson, Jordan, and Hurlbert (2016) measured illumination-discrimination thresholds in dichromats using stimuli similar to those we analyze. For protanopes (corresponding to our MS mosaic) there was a relative elevation of human thresholds in the red and green illumination-change directions, similar to what we see for the computational-observer thresholds. For deuteranopes (corresponding to our LS mosaic), however, thresholds in the red and green illumination-change directions were also elevated, which is inconsistent with the effect seen with the computational observer.

Alvaro et al. (2017) also measured illumination-discrimination threshold in dichromats, but using stimuli derived from natural hyperspectral images and only along blue and yellow illuminant-change directions. They found that thresholds were slightly elevated overall for dichromats relative to trichromats, with no statistically significant interaction between observer type and illuminant-change direction. Given that the changes in relative computational-observer thresholds across LMS, LS, and MS mosaics for the blue and yellow directions are not large, the lack of interaction may be consistent with the computational-observer results.

Modeling illumination discrimination: Modeled Experiment 2

Radonjić et al. (2016) reported results from an illumination-discrimination experiment similar to the one described above, but where across conditions the ensemble of surfaces in the stimulus scene was varied. They obtained illumination-discrimination thresholds for three separate scene surface ensembles, which they labeled neutral, yellowish-green, and reddish-blue (based on the appearance of the corresponding images). The basic methodology of their experiment was the same as for Experiment 1 above, except for the variation of surface ensemble and a slightly different geometric structure of the stimulus scene. Note that in the experiment, the ensemble of surfaces was held constant within each trial so that the question investigated was how the surface ensemble affected illuminant discrimination, not how well observers could tell the difference between changes in illumination and changes in surface ensemble (see Craven & Foster, 1992 for a study of the latter sort). In addition, eye fixations were not measured. The detailed methods are provided in the published report (Radonjić et al., 2016). Figure 9 shows the illumination-discrimination thresholds for each of the surface ensemble conditions (panels D

through F). Illumination-discrimination thresholds depend strongly on the ensemble of surfaces in the scene. Radonjić et al. conjectured that the threshold variation was driven by changes in the information available in the image to make the discriminations. We investigate that conjecture here.

We computed computational-observer performance using the methods described above for the experimental stimuli of Radonjić et al. (2016). Figure 9 shows the results of the analysis. The top row (panels A through C) shows the computational-observer thresholds as a function of the noise factor for the neutral, reddish-blue, and yellowish-green scene conditions. The bottom row (panels D through F) shows the fits of the computational-observer to the psychophysical data. As above, the computational-observer noise factor was fit separately for each observer before averaging across observers. A single noise factor was found for each observer, held fixed across the three surface ensembles and four illumination-change directions. Since eye fixation data were not available, performance was aggregated over all stimulus patches with equal weight.

Contrary to the conjecture of Radonjić et al. (2016), there is little variation in the computational-observer thresholds across the change in surface ensemble, relative to the variation in human psychophysical thresholds. The small effects seen are an overall decrease in predicted thresholds for the yellowish-green ensemble (panel F vs. panels D and E) and a relative decrease in threshold for the red illuminant-change direction for the reddish-blue ensemble (panel E vs. panels D and F). These two changes do not capture the two striking effects in the psychophysical data, a large decrease in red illuminant-change direction threshold between the neutral/yellowish-green and the reddish-blue ensemble, and a large increase in the blue illuminant-change direction threshold between the neutral/reddish-blue and the yellowish-green ensemble.

Discussion

We implemented a computational observer that performs the illumination-discrimination task. The computational observer analysis incorporates well-established properties of early human vision, particularly wavelength-dependent optical blur and sampling by the interleaved LMS cone mosaic, and studies the implications of these properties for human performance. Here, the computational observer has access to the representation at the cone mosaic, with predictions of performance based on a learned linear SVM classifier. By setting up a detailed model of visual processes up to and including isomerization of cone photopigment, we ask about the degree to which
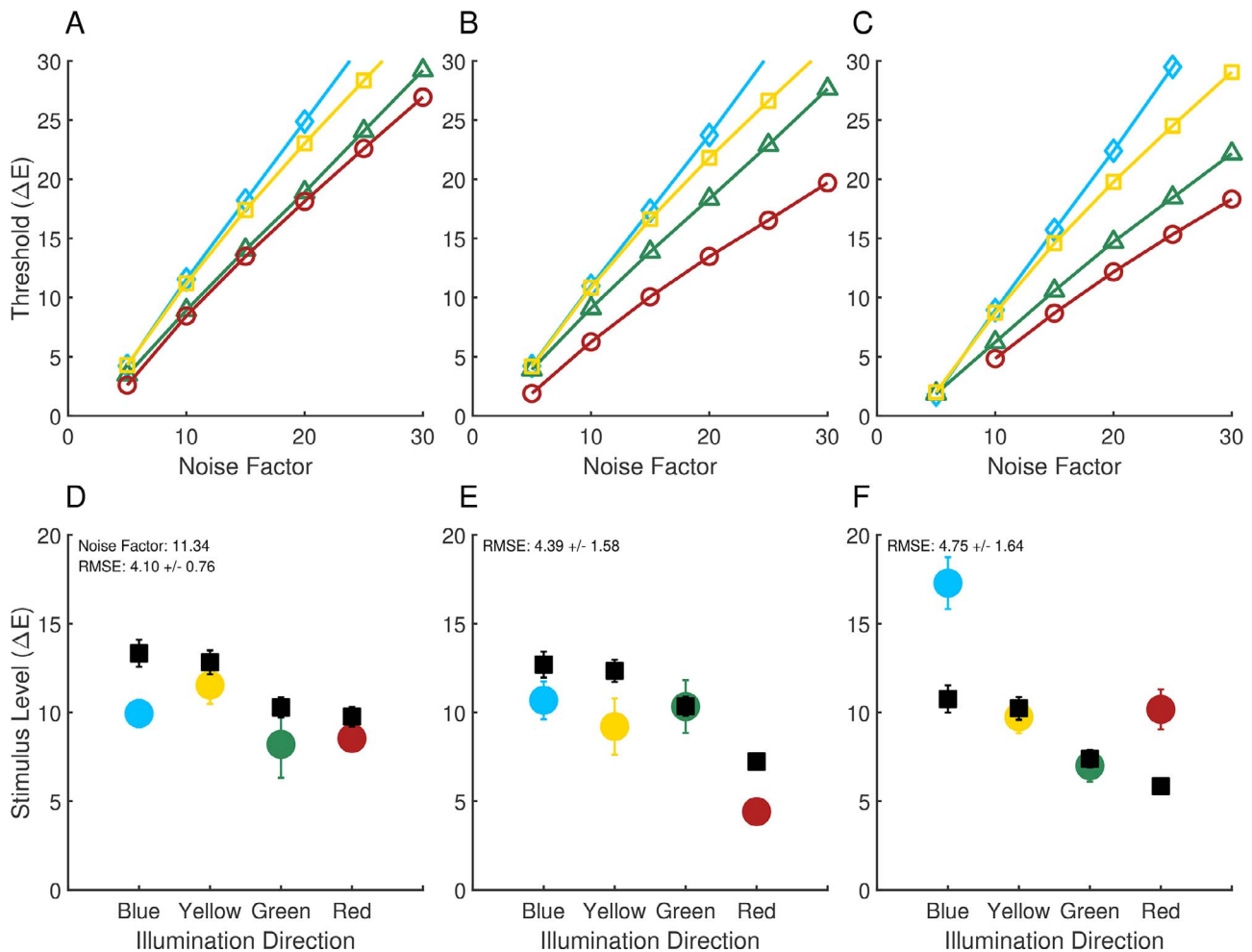
Figure 9. Comparison of computational and psychophysical performance: Experiment 2. Computational-observer thresholds as a function of noise factor for the neutral (A), reddish-blue (B) and yellowish-green condition (C). Computational-observer thresholds fit to human experimental data are also shown for the neutral (panel D), reddish-blue (E) and yellowish-green condition (F). Computational-observer performance was aggregated over all stimulus patches (without applying fixation-based weighting, as we do not have fixation data for these measurements). Fits for panels (D–F) share the same noise factor (11.34). The data in the neutral condition (panel A) represent an unusual case where human thresholds were not highest in the blue illumination-change direction. Experimental data originally reported in Radonjić et al. (2016).

understanding these processes is sufficient to predict human performance. Across the stimulus conditions we studied, thresholds determined from the computational observer's performance do not predict either the absolute or the relative performance of human observers across illumination-change directions and across variation in the ensemble of surfaces in the scene. Absolute deviations are not unexpected, as earlier studies of ideal observer models at the level of the isomerizations generally reveal an overall difference in efficiency between human and ideal observers (Geisler, 1989). It is the relative deviations that are of most interest. These deviations are particularly clear for the case where the ensemble of surfaces in the scene was varied (comparison of computational observer to data reported by Radonjić et al., 2016). Whereas human performance showed a large effect of the surface

ensemble, the computational observer's performance stayed relatively consistent across the ensembles. In addition, the computational observer did not account for the elevated thresholds (expressed as CIELUV ΔE) typically seen for illuminant changes in a blue direction. Thus our results implicate postisomerization processing as an important determinant of not only the absolute level of sensitivity to illumination changes, but also the observed patterns of illumination-discrimination thresholds.

An attractive feature of the computational observer analysis is that we can use it to understand how variation in properties of early vision affects performance in well-specified visual tasks. An example of this is our exploration of the effects of changing the retinal complement of spectral cone types on illumination discrimination (Figure 8). As another example, we

Ding et al.

explored the effect of optical blur on illumination discrimination. Here we found little change in computational observer thresholds when we eliminated optical blur from the computational pipeline (results not shown). This is not surprising, given the relatively large spatial scale of the stimulus patches relative to the point spread function of the human eye.

Limitations

Our computational observer incorporates a number of simplifications. One of these is to implement the observer on 1.1 × 1.1 patches evenly spaced across the stimulus, instead of simulating a larger mosaic with varying cone density that samples the entire stimulus. This choice was made for computational efficiency, particularly with respect to learning the classifier. One long-term goal of our project is to improve the computational efficiency of our methods so that we can simulate larger cone mosaics.

We recognize that a complementary approach is to build simplified calculations that contain only the information necessary for a specific stimulus and task. For example, it seems likely that illumination-discrimination performance for our stimuli may be predicted from the low spatial frequency information in the retinal image. In that case a simplified model, more efficient than the full model, might be appropriate. A special case might replace the fine spatial sampling of a realistic cone mosaic with a blurred and subsampled representation that enables efficient computation.

Although implementing special case calculations is possible, our overall goal is to develop and make available a quantitative image-computable model of early visual processing that may be used to analyze performance on many visual tasks for a wide range of visual stimuli. For that reason, at this juncture, we think it is important for us to focus on extending the model to incorporate visual processing that occurs after the isomerization of photopigment by light (see subsection Postisomerization Processes below). We do acknowledge that special-case calculations derived from the more general full model also represent an interesting future direction.

For Modeled Experiment 1, we compared computational-observer thresholds when we weighted all stimulus patches equally and when we weighted them according to observers' measured eye fixations. The two analyses yielded similar patterns of thresholds. We were not able to make this comparison for Experiment 2 nor for hypothetical observers with di- and monochromatic cone mosaics, because we do not have fixation data available for those cases. There are cases where which patch the computational observer is trained and evaluated on has a large effect on

performance (Figure 5A, B). This suggests that the information about where observers look when they perform psychophysical tasks is useful for computational observer modeling.

Other simplifications of our approach were that we assumed that observers used the information from only a single patch across the duration of each trial, and that we restricted our modeling to two comparison intervals, while excluding the reference interval. Building computational-observer performance models that take into account the trial-by-trial sequence of eye fixation locations across the entire trial is an interesting extension for future research. As with increasing the size of the modeled cone mosaic, this extension would also require an increase in computational resources.

Postisomerization processes

Although it is of fundamental interest to understand how the information available in the responses of the cone mosaic varies across experimental conditions within a psychophysical study, it should not be surprising that there are cases where the rest of the visual system plays a role in shaping performance. Indeed, two well-known processes that occur after photopigment isomerization are likely to influence performance on the illumination-discrimination task. First, regulation of sensitivity (aka adaptation) that begins with the conversion of isomerization rate to photocurrent within the cones has the potential to affect postisomerization information (Hood & Finkelstein, 1986; Stockman & Brainard, 2010; Angueyra & Rieke, 2013). As light level increases, this adaptation limits sensitivity in a manner that dominates the effects of Poisson-distributed noise in the photopigment isomerizations, producing Weber's Law rather than square root law behavior in the dependence of threshold on light level. The luminance at which cone sensitivity regulation is manifest in psychophysical data depends on the spatial and temporal properties of the stimulus (Hood & Finkelstein, 1986), but it may well account for some of the differences between the performance of the computational observer presented here and that of human observers.

Second, signals from the three classes of cones are recombined by postisomerization processing into luminance and cone-opponent channels (Shevell & Martin, 2017). To the extent that noise following this recombination limits discrimination, these processes can have a major effect on the relative sensitivity of the visual system to stimulus changes in different directions in color space (for a review, see Stockman & Brainard, 2010). Indeed, it has been shown that cone-opponent processing plays an important role in shaping sensitivity to modulations in different color directions in

psychophysical tasks that involve simple colored stimuli, such as spots or Gabor patches seen against a spatially uniform background (Wandell, 1995; Stockman & Brainard, 2010).

Our current implementation of the computational observer uses additive zero-mean Gaussian noise as a proxy for all of postisomerization vision. This noise does not model stimulus-specific effects, nor should it be expected to capture all aspects of visual mechanisms beyond the site of photopigment isomerization. Rather, it was included because real observer thresholds are considerably higher than those of the computational observer, even when we consider only $1.1° \times 1.1°$ patches of retina. We are eager to extend our computational-observer models to incorporate both the conversion of isomerization rate to photocurrent and the recombination of signals from different classes of cones by retinal ganglion cells, along the lines of the sequential ideal observer analysis outlined by Geisler (1989). A computational observer that models multiple stages of neural information processing in the retina will likely provide a better account of the experimental data with less need for an omnibus parameter to reduce overall efficiency.

Keywords: color vision, illumination discrimination, ideal observer, computational observer

Footnotes

[1] Metamers for natural daylights were used to match spectra that could be produced by an apparatus designed to study illumination discrimination for real illuminated surfaces. Radonjić et al. (2016) present a comparison of illumination-discrimination performance for scenes consisting of real illuminated surfaces and graphics-based renderings. Finlayson, Mackiewicz, Hurlbert, Pearce, and Crichton (2014; see also Pearce et al., 2014) describe in some detail the daylight metamers themselves.

[2] The value of 1 ΔE is a nominal step size. The actual step sizes varied across illuminant directions and modeled experiments. The actual step sizes were used in all analyses. See Radonjić et al. (2016) and Radonjić et al. (2018) for details.

[3] Stimulus images in Modeled Experiment 1 were viewed from 68.3 cm and subtended 20.0° by 16.7° of visual angle. Images in Modeled Experiment 2 were viewed from 76.4 cm and subtended 18.6° by 17.3° of visual angle. The corresponding scene specifications in ISETBio used the same viewing distances and sizes as those in the experiments.

[4] Because rendering is a stochastic process, it is possible to train a classifier to distinguish between two separate renderings of the same scene, even though the differences are imperceptible to a human observer. In Modeled Experiment 1, we used 30 separate renderings of the target image and drew randomly from these on each psychophysical trial to generate each training/test set instance. Our modeling followed this same procedure. In Modeled Experiment 2 described below, only one target image was used in the psychophysics. In the modeling, we generated seven versions of that same target image and drew randomly from these to generate each training/test set instance. The two draws for each trial (one for target interval and the other for one of the comparison intervals) were made without replacement.

References

Alvaro, L., Linhares, J. M. M., Moreira, H., Lillo, J., & Nascimento, S. M. C. (2017). Robust colour constancy in red-green dichromats. *PLoS One*, *12*(6):e0180310.

Angueyra, J. M., & Rieke, F. (2013). Origin and effect of phototransduction noise in primate cone photoreceptors. *Nature Neuroscience*, *16*(11), 1692–1700.

Aston, S., Radonjić, A., Brainard, D. H., & Hurlbert, A. C. (2019). Illumination discrimination for chromatically biased illuminations: Implications for colour constancy. *Journal of Vision*, 19(30):15, 1–23, https://doi.org/10.1167/19.3.15. [PubMed] [Article]

Aston, S., Turner, J., Le Couteur Bisson, T., Jordan, G., & Hurlbert, A. C. (2016). Better colour constancy or worse discrimination? Illumination discrimination in colour anomalous observers. *Perception (Supplement)*, 45(2), Abstract 3P035.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research*, 27(11), 1915–1924.

Barlow, H. B. (1962). Measurements of the quantum efficiency of discrimination in human scotopic vision. *Journal of Physiology*, 160(1), 169–188.

Beck, J. (1959). Stimulus correlates for the judged illumination of a surface. *Journal of Experimental Psychology*, 58(4), 267–274.

Bone, R. A., Landrum, J. T., & Cains, A. (1992). Optical density spectra of the macular pigment *in vivo* and *in vitro*. *Vision Research*, 32(1), 105–110.

Brainard, D. H. (2015). Color and the cone mosaic. *Annual Review of Vision Science*, 1, 519–546.

CIE. (2004). *Colorimetry, third edition* (No. 15). Vienna: Bureau Central de la CIE.

CIE. (2006). *Fundamental chromaticity diagram with physiological axes—Part 1* (Report No. 170–171). Vienna: Central Bureau of the Commission Internationale de l' Éclairage.

Cottaris, N. P., Jiang, H., Ding, X., Wandell, B. A., & Brainard, D. H. (2018). A computational observer model of spatial contrast sensitivity: Effects of wavefront-based optics, cone mosaic structure, and inference engine. *Journal of Vision*, 19(4):8, 1–27, https://doi.org/10.1167/19.4.8. [PubMed] [Article]

Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research*, 32(7), 1359–1366.

Curcio, C. A., Allen, K. A., Sloan, K. R., Lerea, C. L., Hurley, J. B., Klock, I. B., & Milam, A. H. (1991). Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *Journal of Comparative Neurology*, 312(4), 610–624.

Farrell, J. E., Jiang, H., Winawer, J., Brainard, D. H., & Wandell, B. A. (2014). Distinguished Paper: Modeling visible differences: The computational observer model. *SID Symposium Digest of Technical Papers*, 45, 352–356, https://doi.org/10.1002/j.2168-0159.2014.tb00095.x.

Finlayson, G., Mackiewicz, M., Hurlbert, A., Pearce, B., & Crichton, S. (2014). On calculating metamer sets for spectrally tunable LED illuminators. *Journal Of The Optical Society Of America A*, 31(7), 1577–1587.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, 96(2), 267–314.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, 51(7), 771–781.

Gilchrist, A., & Jacobsen, A. (1984). Perception of lightness and illumination in a world of one reflectance. *Perception*, 13, 5–19.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hecht, S., Schlaer, S., & Pirenne, M. H. (1942). Energy, quanta and vision. *Journal of the Optical Society of America*, 38(6), 196–208.

Hernandez-Andres, J., Romero, J., Nieves, J. L., & Lee, R. L. (2001). Color and spectral analysis of daylight in southern Europe. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 18(6), 1325–1335.

Hofer, H. J., Carroll, J., Neitz, J., Neitz, M., & Williams, D. R. (2005). Organization of the human trichromatic cone mosaic. *Journal of Neuroscience*, 25(42), 9669–9679.

Hood, D. C., & Finkelstein, M. A. (1986). Sensitivity to light. In K. Boff, L. Kaufman & J. Thomas (Eds.), *Handbook of Perception and Human Performance* (*Vol. 1*, pp. 5-1–5-66). New York: Wiley.

Hurlbert, A. (1989). *Cues to the color of the illuminant.* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Jiang, H., Cottaris, N. P., Golden, J., Brainard, D., Farrell, J. E., & Wandell, B. A. (2017). *Simulating retinal encoding: Factors influencing Vernier acuity.* Paper presented at Electronic Imaging 2017, Burlingame, CA.

Kardos, L. (1928). Dingfarbenwahrnehmung und Duplizitätstheorie. [The perception of object colors and duplicity theory] *Zeitschrift für Psychologie*, 108, 240–314.

Koenig, D., & Hofer, H. (2011). The absolute threshold of cone vision. *Journal of Vision*, 11(1):21, 1–24, https://doi.org/10.1167/11.1.21. [PubMed] [Article]

Kolb, H. Photoreceptors. *Webvision*. Retrieved from http://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/photoreceptors/

Kozaki, A., & Noguchi, K. (1976). The relationship between perceived surface-lightness and perceived illumination. *Psychological Research*, 39, 1–16.

Lee, T. Y., & Brainard, D. H. (2011). Detection of changes in luminance distributions. *Journal of Vision*, 10(13):14, 1–16, https://doi.org/10.1167/10.13.14. [PubMed] [Article]

Logvinenko, A., & Menshikova, G. (1994). Trade-off between achromatic colour and perceived illumination as revealed by the use of pseudoscopic inversion of apparent depth. *Perception*, 23(9), 1007–1023.

Logvinenko, A. D., & Maloney, L. T. (2006). The

proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching. *Perception & Psychophysics*, *68*(1), 76–83.

Lopez, H., Murray, H. L., & Goodenough, D. J. (1992). Objective analysis of ultrasound images by use of a computational observer. *IEEE Transactions on Medical Imaging*, *11*(4), 496–506.

Lucassen, M. P., Gevers, T., Gijsenij, A., & Dekker, N. (2013). Effects of chromatic image statistics on illumination induced color differences. *Journal Of The Optical Society Of America A*, *30*(9), 1871–1884.

Manning, C. D., Raghavean, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge, UK: Cambridge University Press.

Nascimento, S. M., Amano, K., & Foster, D. H. (2016). Spatial distributions of local illumination color in natural scenes. *Vision Research*, *120*, 39–44.

Noguchi, K., & Kozaki, A. (1985). Perceptual scission of surface lightness and illumination: An examination of the Gelb effect. *Psychological Research*, *47*(1), 19–25.

Oyama, T. (1968). Stimulus determinants of brightness constancy and the perception of illumination. *Japanese Psychological Research*, *10*(3), 146–155.

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLoS One 9*(2):e87989.

Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox. *Frontiers in Psychology*, *9*: 1250, https://doi.org/10.3389/fpsyg.2018.01250.

Radonjić, A., Ding, X., Krieger, A., Aston, S., Hurlbert, A. C., & Brainard, D. H. (2018). Illumination discrimination in the absence of a fixed surface-reflectance layout. *Journal of Vision*, *18*(5):11, 1–27, https://doi.org/10.1167/18.5.11. [PubMed] [Article]

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., . . . Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision*, *16*(11):2, 1–18, https://doi.org/10.1167/16.11.2. [PubMed] [Article]

Rodieck, R. W. (1998). *The first steps in seeing*. Sunderland, MA: Sinauer.

Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: A direct test of the illumination estimation hypothesis. *Psychological Science*, *13*, 142–149.

Sekiguchi, N., Williams, D. R., & Brainard, D. H. (1993). Efficiency in detection of isoluminant and isochromatic interference fringes. *Journal of the Optical Society of America A*, *10*(10), 2118–2133.

Shevell, S. K., & Martin, P. R. (2017). Color opponency: Tutorial. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *34*(7), 1099–1108.

Spitschan, M., Aguirre, G. K., Brainard, D. H., & Sweeney, A. M. (2016). Variation of outdoor illumination as a function of solar elevation and light pollution. *Scientific Reports*, *6*:26756.

Stockman, A., & Brainard, D. H. (2010). Color vision mechanisms. In M. Bass et al. (Eds.), *The Optical Society of America handbook of optics* (3rd ed., *Vol. III*: *Vision and vision optics*, pp. 11.1–11.104). New York, NY: McGraw Hill.

Stockman, A., Sharpe, L. T., & Fach, C. C. (1999). The spectral sensitivity of the human short-wavelength cones. *Vision Research*, *39*(17), 2901–2927.

Thibos, L. N., Hong, X., Bradley, A., & Cheng, X. (2002). Statistical variation of aberration structure and image quality in a normal population of healthy eyes. *Journal of the Optical Society of America A*, *19*(12), 2329–2348.

Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer.

Watson, A. B., & Yellott, J. I. (2012). A unified formula for light-adapted pupil size. *Journal of Vision*, *12*(10):12, 1–16, https://doi.org/10.1167/12.10.12. [PubMed] [Article]

Weiss, D., & Gegenfurtner, K. (2016). A comparison between illuminant discrimination and chromatic detection. *Perception*, *45*(S2), 214.

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *The British Journal of Mathematical and Statistical Psychology*, *18*, 1–10.

Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae* (2nd ed.). New York, NY: John Wiley & Sons.