

A Bayesian model of lightness perception that incorporates spatial variation in the illumination

Sarah R. Allred

Department of Psychology, Rutgers,
The State University of New Jersey, Camden, NJ, USA



David H. Brainard

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



The lightness of a test stimulus depends in a complex manner on the context in which it is viewed. To predict lightness, it is necessary to leverage measurements of a feasible number of contextual configurations into predictions for a wider range of configurations. Here we pursue this goal, using the idea that lightness results from the visual system's attempt to provide stable information about object surface reflectance. We develop a Bayesian algorithm that estimates both illumination and reflectance from image luminance, and link perceived lightness to the algorithm's estimates of surface reflectance. The algorithm resolves ambiguity in the image through the application of priors that specify what illumination and surface reflectances are likely to occur in viewed scenes. The prior distributions were chosen to allow spatial variation in both illumination and surface reflectance. To evaluate our model, we compared its predictions to a data set of judgments of perceived lightness of test patches embedded in achromatic checkerboards (Allred, Radonjić, Gilchrist, & Brainard, 2012). The checkerboard stimuli incorporated the large variation in luminance that is a pervasive feature of natural scenes. In addition, the luminance profile of the checks both near to and remote from the central test patches was systematically manipulated. The manipulations provided a simplified version of spatial variation in illumination. The model can account for effects of overall changes in image luminance and the dependence of such changes on spatial location as well as some but not all of the more detailed features of the data.

Introduction

For the perceived color of an object to be useful in recognition and discrimination, this color should remain relatively stable across changes in the environment in which the object is viewed. Such stability is

called color constancy. Achieving color constancy represents a challenge for the visual system because the light reflected from an object depends both on the object's intrinsic reflectance spectrum and on the spectral power distribution of the illuminant. Everyday experience suggests that human color vision exhibits reasonable color constancy. We can use color, for example, to reliably distinguish a lemon from a lime under both indoor and outdoor lighting. This introspective conclusion is supported by the empirical literature, in which substantial color constancy is found when objects are viewed in reasonably realistic environments (for recent reviews see Brainard & Radonjić, in press; Foster, 2011; Shevell & Kingdom, 2008; Smithson, 2005).

It is not currently understood how the visual system resolves surface-illuminant ambiguity to achieve constancy. One approach to this question is to ask how an ideal observer might do so. This approach has long been appealing (Helmholtz, 1910), but only recently has it been possible to use Bayesian statistical theory and related computational approaches to develop quantitative algorithms whose performance may be linked to and compared with human performance (Knill & Richards, 1996). Bayesian methods provide a principled method using the statistical regularities of natural scenes to resolve the ambiguity inherent in image data.

In the area of constancy, Bayesian approaches have had success in modeling human performance for restricted classes of scenes in which a collection of matte surfaces are illuminated by a single diffuse illuminant (Brainard, 2009; Brainard et al., 2006). Most natural scenes, however, are not diffusely illuminated, and there are empirical phenomena that cannot be accounted for by models based on Bayesian algorithms that assume a single scene illuminant. Indeed, any scene in which the mapping between the retinal image and the perception of color depends on spatial location is

Citation: Allred, S. R., & Brainard, D. H. (2013). A Bayesian model of lightness perception that incorporates spatial variation in the illumination. *Journal of Vision*, 13(7):18, 1–18, <http://www.journalofvision.org/content/13/7/18>, doi:10.1167/13.7.18.

inconsistent with predictions derived from Bayesian algorithms that assume a single spatially homogeneous scene illuminant (for discussion, see Brainard & Maloney, 2011). An important challenge for the Bayesian program then is whether the approach can be generalized successfully to account for performance measured with stimuli that elicit geometric effects. Doing so requires developing a Bayesian algorithm that allows for spatial variation in the scene illumination and then linking the performance of the algorithm to human performance for an appropriate ensemble of experimental stimuli.

Given the astronomical numbers of images that could be presented to a subject and processed by an algorithm, it does not seem wise to step from the study of scenes with a single diffuse illuminant to the full complexity of natural images. Thus we must make some choice that restricts the stimulus ensemble that will be studied. Our first restriction in the present work is from the full color case to achromatic stimuli that vary only in luminance. In this case, the corresponding perceptual dimension is lightness. Within this domain, recent work has helped organize a wide range of empirical lightness phenomena. In particular, both Gilchrist and Adelson theorize that it is useful to think about lightness perception in terms of frameworks (Gilchrist, 2006; Gilchrist et al., 1999) or atmospheres (Adelson, 2000). They suggest that complex scenes can be characterized as being composed of elemental regions that are approximately uniformly illuminated. A successful theory would explain how the visual system segments the image into such regions, how each region is processed, and the degree to which the processing of separate regions interacts. The framework or atmosphere conceptualization thus suggests that an important step in extending the Bayesian approach is to generalize to scenes that, while still simple, are naturally separated into multiple such frameworks or atmospheres and to ask whether performance for such scenes may be modeled with an appropriately elaborated Bayesian algorithm. Here we proceed along these lines.

We consider scenes consisting of achromatic checkerboards, for which we have previously reported an extensive set of psychophysical lightness data (Allred et al., 2012). The scenes have the feature that large differences in image luminance between subregions of the checkerboards provide cues that support segmentation of the checkerboards into separate frameworks or atmospheres. We develop a Bayesian algorithm that estimates illumination and reflectance from image luminance for the checkerboard scenes, and elaborate the algorithm into a model of psychophysical performance by linking its output to judgments of perceived lightness. The algorithm resolves surface-illuminant ambiguity via specified priors over illumination and

surface reflectance. We chose the parametric form of the priors to express, within the checkerboard scene domain, intuitions about the properties of naturally occurring illumination and surface reflectances. Given the parametric form of the priors, the model's predictions are determined by the particular parameters that specify the priors. We fit the model to the psychophysical data by using numerical parameter search to find the prior parameters that resulted in the best model fit, and we evaluate and discuss the resultant account of the data.

Methods

Algorithm

We developed a Bayesian algorithm that estimates illumination and surface reflectance from image data, for a restricted class of scenes. The scenes consisted of regular achromatic checkerboards. Thus each surface in the checkerboard was specified by its location and its scalar reflectance, $r_{i,j}$, where i and j denote the row and column location of the square, respectively. In our experiments, we employed 5×5 checkerboards, so that i and j ranged between 1 and 5. The entire checkerboard of surfaces was described by the column vector \vec{r} whose entries are the $r_{i,j}$ in raster order.

We allowed the illumination to vary spatially, but for simplicity required that it be constant over each checkerboard surface. Thus the illumination was described by the luminance incident on each surface in the checkerboard, $e_{i,j}$. This was summarized for the entire scene by the column vector \vec{e} .

The vector describing the scene, which we refer to as the world vector, was taken as the concatenation of the vectors \vec{r} and \vec{e} , so that $\vec{w} = [\vec{r}, \vec{e}]$. Although the scenes we considered were greatly simplified relative to those encountered in natural viewing, they embodied two key features. These were the fundamental illuminant-surface ambiguity that is characteristic of the problem of color constancy and the fact that both reflectance and illumination could vary across locations within a scene.

Given the visual world of achromatic checkerboard scenes, the sensory image was given by the reflected luminance $l_{i,j}$ at each checkerboard location. This was described by a column vector \vec{l} . The reflected luminance at each location was taken as the product of the corresponding illuminant and reflectance: $l_{i,j} = e_{i,j} r_{i,j}$. The algorithm's task was to estimate \vec{w} from \vec{l} . This is clearly an underdetermined problem, since \vec{w} has twice as many entries as \vec{l} . To formulate constraints on the solution and develop an algorithm to find \vec{w} from \vec{l} , we

employed Bayesian decision theory (Berger, 1985; Lee, 1989).

There are three key ingredients required to develop a Bayesian algorithm. The first is the *likelihood*. This expresses the relationship between the representation of the visual world (\vec{w}) and the observed data (\vec{l}) as a probability distribution $P(\vec{l}|\vec{w})$. The likelihood characterizes the probability with which a set of luminance values \vec{l} would be observed if the world actually contained the surfaces and illuminants described by \vec{w} . For computer vision applications, we can think of the likelihood as a probabilistic way to describe the imaging process. In general, calculation of the likelihood involves incorporation of processes that perturb or add noise to an incident signal (e.g., optics of the eye and photon noise). Here, however, we assumed that the encoded luminance was noise-free, so that:

$$P(\vec{l}|\vec{w}) = \begin{cases} d, & \text{if } l_{i,j} = r_{i,j} e_{i,j} \text{ for all } i,j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where d is a constant. Using a noise-free likelihood means that algorithm performance is governed by how the prior, described in following text, resolves the ambiguity about reflectance introduced by uncertainty about the illumination.

The second ingredient for a Bayesian algorithm is the *prior*. This captures statistical regularities of the visual world as a probability distribution $P(\vec{w})$. We chose a prior that expressed several assumptions about the visual world. First, the surfaces in a scene are drawn independently from the illumination. Thus, $P(\vec{w}) = P(\vec{r}, \vec{e}) = P(\vec{r})P(\vec{e})$.

Second, we assumed that the surface reflectances within a checkerboard were independently and identically distributed, so that $P(\vec{r}) = \prod_{i,j} P(r_{i,j})$. We took the reflectance distribution at each image location to be a beta distribution

$$P(r_{i,j}) = B(r_{i,j}; \alpha_{surface}, \beta_{surface}). \quad (2)$$

The beta is defined over the range 0 to 1. The relative probability of surfaces of different reflectance is adjusted by the parameters $\alpha_{surface}$ and $\beta_{surface}$.

Third, we assumed that the illuminant varied more slowly across the array than the surface reflectances. This idea, which seems intuitively reasonable, has been used in previous surface/illuminant estimation algorithms that allowed the illuminant to vary across spatial locations (Funt & Drew, 1988; Land & McCann, 1971). To capture this in the prior distribution, we took the illuminant prior to be a multivariate lognormal

$$P(\vec{e}) = \ln N(\vec{e}; \vec{\mu}_{illum}, \mathbf{K}_{illum}). \quad (3)$$

The lognormal is defined over positive values and has a long positive tail. This allows the prior to account for a wide range of illuminant intensities. The mean

illuminant intensity is determined by the parameter vector $\vec{\mu}_{illum}$, which provided the mean value at each location. We chose a spatially uniform mean, so that each entry of $\vec{\mu}_{illum}$ was given by a single parameter μ_{illum} . The lognormal also has a covariance matrix \mathbf{K}_{illum} , which allowed us to specify that illuminant intensities at neighboring locations are correlated. Such specification captures the assumption that the illuminant varies slowly over space. How slowly the illuminant varies is determined by the exact structure of the covariance matrix. Indeed, \mathbf{K}_{illum} was constructed to represent a first-order Markov field, so that the correlational structure was controlled by a single parameter ρ_{illum} . Let the variance of the illuminant intensity at each location be the same and be given by σ_{illum}^2 . Then the covariance between illuminant intensities at locations $[i, j]$ and $[k, l]$ was given by $\sigma_{illum}^2 \rho_{illum}^{|i-k|+|j-l|}$.

The likelihood and prior were combined using Bayes' rule to calculate the *posterior*:

$$P(\vec{w}|\vec{l}) = c P(\vec{l}|\vec{w}) P(\vec{w}) \quad (4)$$

where c is a normalizing constant. The posterior combines the likelihood and the prior and describes the probability of any visual world given the observed luminance values.

The third ingredient for a Bayesian algorithm is to specify a rule for choosing an actual estimate from the posterior. Here we chose the \vec{w} that maximized the posterior. To find this \vec{w} for a set of luminances \vec{l} and a set of prior parameters $[\alpha_{surface}, \beta_{surface}, \mu_{illum}, \sigma_{illum}^2, \rho_{illum}]$ we used numerical search as implemented by the `fmincon` function of MATLAB (Mathworks, Natick, MA). Because we assumed a noise-free likelihood, it was sufficient to search only over the space of illuminant vectors \vec{e} , since each choice of \vec{e} allowed computation of the \vec{r} that was consistent with it and the observed luminances \vec{l} . Thus our parameter search was over a 25-dimensional space. We bounded the searched illuminant intensities to lie between 0.001 and 30.

It was also critical to start the search with reasonable initial guesses as to the estimates. To produce a set of such guesses, we took 2,000 draws from the prior distribution, and found a set of n -dimensional linear models for the space of illuminants (where n took values of [2, 4, 6, 9, 10, 12, 14]). We searched over illuminants within each of these linear models in order of increasing dimension, using the result of the preceding search as the initial guess for the next. The estimate of \vec{e} that resulted in the highest posterior from this preliminary optimization was used as the initial guess for the full dimensional problem. For a subset of conditions, we investigated the sensitivity of our search procedures to the initial guess. With some guesses, the `fmincon` search simply returned the initial guess. We detected and rejected these cases. For the other initial

guesses, the returned solution was independent of the initial guess. This check provides some assurance that the returned solutions approximate global maxima of the posterior, although we cannot know this with certainty.

For a subset of conditions, we also verified that searching across \vec{r} did not yield different solutions than searching across \vec{e} .

In summary then, for a given set of parameters $[\alpha_{surface}, \beta_{surface}, \mu_{illum}, \alpha_{illum}^2, \rho_{illum}]$ and a set of luminance values \vec{l} , our algorithm estimates the reflectance and illuminant values that are most likely. That is, our estimate is the \vec{w} that maximizes $P(\vec{w}|\vec{l})$.

Psychophysics

The methods used to collect the psychophysical data, as well as the data themselves, are described in detail in Allred et al. (2012) and summarized here. Briefly, seven observers looked through an aperture into a rectangular enclosure, at the end of which they viewed an achromatic 25-square checkerboard presented on a custom-built high-dynamic range display (see Radonjić, Allred, Gilchrist, & Brainard, 2011 for display specifications). Observers were asked to judge the lightness of the center square (test patch) by matching it to one of a series of Munsell papers that ranged from 2.0 (black) to 9.5 (white) in 0.5-unit steps.

The test patch (center square) took on 24 distinct luminance values, ranging from 0.096 cd/m² to 211 cd/m². The smallest value was the minimum luminance value of the high-dynamic range display and should be considered approximate. The remainder of the test patches were chosen in equal log steps between 0.24 cd/m² and the maximum luminance of the display 211 cd/m². The patches had CIE xy chromaticity (0.43, 0.40). The same 24 test patches were judged within nine separate checkerboard contexts (Figure 1).

A standard checkerboard context was created by taking 24 luminance values between 0.11 and 211 cd/m² (contrast ratio 1,878:1) that were equidistant in logarithmic units. These 24 luminance values were assigned to a 5 × 5 checkerboard surrounding the center test square. To assign luminance values to squares, we took random draws of spatial arrangement until neither the brightest nor the darkest luminance were in the inner ring immediately adjacent to the center square. This arrangement was used as the standard context in all experiments; a representation of this standard checkerboard context is shown in Figure 1. The remaining eight test checkerboard contexts were created in the following fashion. We divided the 24 checkerboard squares into inner (eight locations immediately adjacent to the center test square) and outer rings (16 locations surrounding the inner ring).

We created low, standard, and high luminance distributions for inner and outer rings (for details, see Allred et al., 2012). Then we assigned each possible permutation of these rings to the eight test checkerboard contexts (i.e., low inner–low outer checkerboard; low inner–standard outer checkerboard; low inner–high outer checkerboard, etc.). The spatial arrangement of the low and high inner and outer rings in each test checkerboard context preserved the rank order of luminance values in the standard checkerboard context.

Note that the test checkerboard contexts were not constructed to simulate a fixed set of papers under different illuminants; that is, neither inner nor outer ring manipulations were implemented as multiplicative factors of the corresponding luminance values for the standard checkerboard context. Thus, it is not straightforward to interpret the psychophysical data in terms of the degree of constancy they reveal. Rather than asking about constancy per se, we ask whether a model derived from an algorithm designed to achieve constancy can predict the observed psychophysical data.

To proceed, we averaged the luminance values matched to each Munsell paper; the data aggregated thus give, for each Munsell paper, a set of nine luminance values (one for each test checkerboard context) that are perceptually equivalent. By plotting the luminance values for each of the eight test checkerboard contexts against the luminance values for the standard checkerboard context we establish eight context transfer functions (CTFs) that characterize the effect of changing context from the standard checkerboard context to the each of the eight test checkerboard contexts. It is these CTFs in particular that we seek to model.

Using the algorithm to model psychophysical lightness judgments

We applied the Bayesian algorithm to the stimuli used in the psychophysical experiments. For any set of algorithm parameters (priors), we obtained estimates of the illuminant and surface reflectance at each checkerboard location from a specification of the luminance in that checkerboard context. In our previous report (Allred et al., 2012) and in the methods summary above, luminance values are reported in units of candelas per square meter; for the calculations, luminance was specified in normalized units whose range was 0 to 1, with 1 equivalent to the maximum luminance displayed in the experiment.

To compare the algorithm's performance to the psychophysical data, we need to specify a linking hypothesis that connects the algorithm's output to the experimental measurements (see Brainard, Kraft, &

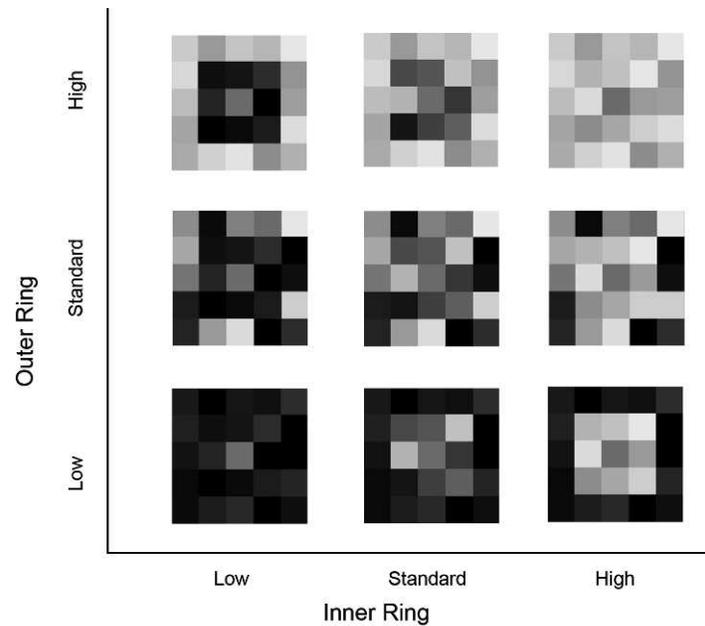


Figure 1. Illustration of the nine experimental checkerboard contexts. Average luminance of inner ring and outer ring were divided into low, standard, and high conditions. The central test patch has the same luminance in all nine checkerboard contexts shown here.

Longé, 2003; Teller, 1984). To do so, we assumed that when the Bayesian algorithm estimated that two luminance values in different contexts [$L_{a(\text{Context } x)}$, $L_{b(\text{Context } y)}$] had the same reflectance (R_z), then these two test luminance values would match in lightness across the context change. This linking hypothesis is based on the general idea that perceived lightness is a perceptual correlate of surface reflectance, but takes into account the fact that reflectance is not explicitly available in the retinal image. The role of the algorithm in the model is to provide a computation that converts proximal luminance to a form that is more plausibly related to perceived lightness.

Given the linking hypothesis above, we computed CTFs for the algorithm that could be compared to the psychophysical CTFs. Indeed, computation of algorithm-based CTFs proceeded in a fashion similar to that used to generate the psychophysical CTFs. The one key difference is that rather than using the matched Munsell papers to establish equivalence across contexts, we used the estimates of surface reflectance returned by the algorithm. Thus the particulars of the computation differed slightly.

First, as described in Methods, we computed algorithm estimates of each of the 216 test-checkerboard luminance combinations viewed by human observers (24 test patches embedded in each of nine checkerboard contexts). Although we computed both illuminant and surface reflectance estimates for all 25 checkerboard locations in each case, the key value that we extracted to compute the CTFs was the estimated surface reflectance at the test location (central test

patch). Then, for each context, we fit estimated test patch reflectance as a function of test luminance with a third-order polynomial. This allowed us to interpolate between the discrete estimated reflectance values. The polynomial functional form was chosen for convenience and has no theoretical significance. Let $R_{estimated} = f_x(L_i)$ represent the interpolated reflectance values, where x represents one of the nine checkerboard contexts and i indicates the 24 test patch values. In the standard context ($x = St$), we evaluated this function for all L_i to obtain a set of reflectance values $[R]_{St}$ that served as the referents for establishing CTFs (much as the Munsell papers did for the psychophysical judgments). To compute a CTF $_x$, we inverted the interpolated function f_x to find the value L that yielded each $[R]_{St}$. Thus, each algorithm-based CTF consists of 24 $[L_{St}, L_x]$ pairs that were taken as perceptually equivalent.

The five parameters $\alpha_{surface}$, $\beta_{surface}$, μ_{illum} , σ_{illum}^2 , and ρ_{illum} control the prior probability and hence drive the algorithm estimates. The parameter values we used for the algorithm were chosen to minimize the average error between algorithm-based CTFs and psychophysical CTFs. To find these values, we used a grid search on the algorithm parameters. We computed algorithm estimates for the 216 test-checkerboard pairs described above for thousands of sets of parameter values. Initial parameters were chosen through visual inspection of model predictions for a variety of simulated scenes. From these initial values, we varied each parameter in coarse steps to determine the best region of parameter space and then sampled this space more finely. Since

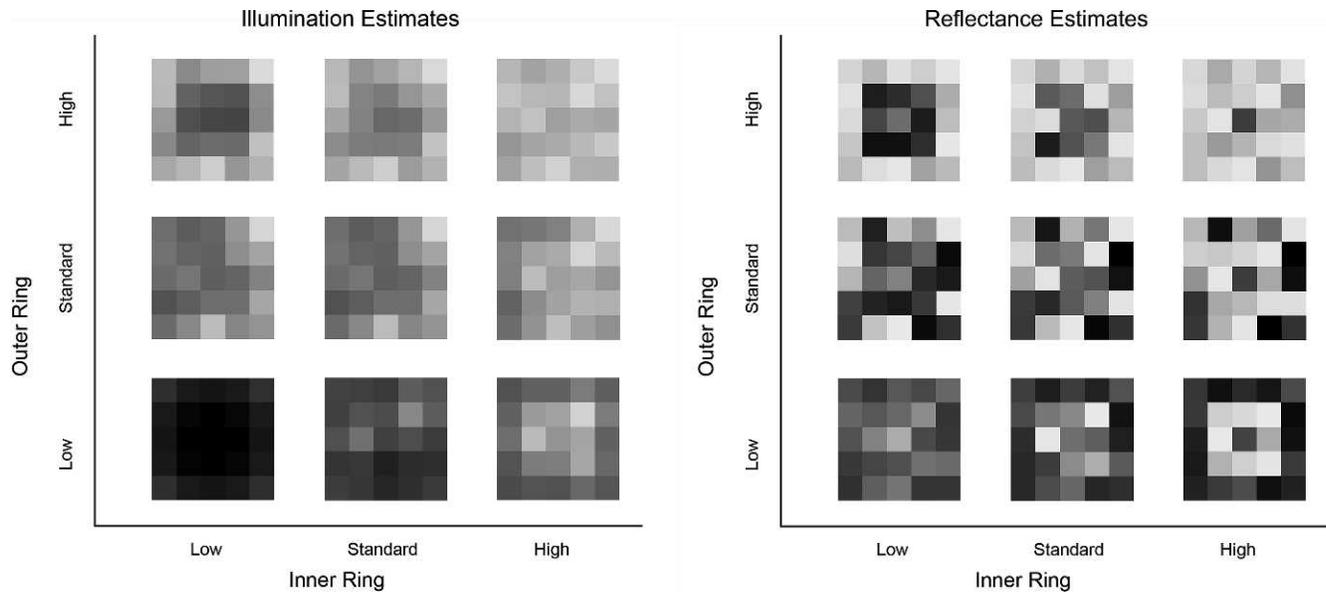


Figure 2. Pictorial interpretation of algorithm estimates of the illuminant (left panel) and reflectance (right panel) of one test patch luminance for each checkerboard context shown in Figure 1. For visualization purposes, the estimated values are scaled and normalized (one factor was applied to all illuminant estimates, another was applied to all reflectance estimates). Checkerboard contexts are grouped by the luminance profile of inner and outer rings, as in Figure 1. The algorithm estimates shown here were obtained using parameters $\alpha_{surface} = 1$, $\beta_{surface} = 2$, $\mu_{illum} = 1$, $\sigma_{illum}^2 = 0.81$, $\rho_{illum} = 0.46$.

our grid search was not exhaustive, it remains possible that a different set of parameter values could fit the data better.

For each set of parameters, we calculated algorithm-based CTFs via the method described above. Algorithm-based CTFs were constructed from 24 $[L_{St}, L_x]$ pairs while the psychophysical CTFs were constructed using the 16 $[L_{St}, L_x]$ defined by the Munsell chips. To directly compare the two sets of CTFs, we interpolated the algorithm-based CTFs to obtain values for each of the 16 psychophysical L_{St} values. We chose final algorithm parameters that minimized the average prediction error in a least-squares sense. We refer to these as the *derived priors* to emphasize that they were obtained by a fit to the psychophysical data, rather than directly from measurements of naturally occurring illuminants and surfaces.

Results

To give intuition about the Bayesian algorithm, we first provide pictorial depictions of algorithm estimates for the nine checkerboard contexts. We then make quantitative comparisons between algorithm estimates and psychophysical measurements of lightness and discuss the aspects of the algorithm that allow it to predict the broad features of the human judgments.

Intuition about the algorithm

Pictorial depictions of algorithm estimates of illumination and reflectance for all nine checkerboard contexts (those in Figure 1) are shown in Figure 2. Inspection of these pictorial depictions provides intuition about the algorithm's behavior.

First, the algorithm estimates that illumination varies systematically across locations, for most of the contexts. When the inner and outer rings of the checkerboard were drawn from very different luminance distributions, such as in the low–high (upper left panel in Figure 2) or high–low (bottom right panel in Figure 2) checkerboard contexts, the algorithm estimated higher spatial variation in the illuminant than when inner and outer rings of the checkerboard were drawn from the same luminance distributions, such as in the low–low (bottom left panel in Figure 2), standard (center panel in Figure 2), or high–high (top right panel in Figure 2) checkerboard contexts. This is consistent with a visual interpretation of the scene as having a shadow at the center of the low-high checkerboard context, a spotlight at the center of high–low checkerboard context, and relatively uniform illumination for the low–low, standard, and high–high checkerboard contexts.

Second, the overall algorithm estimates of the illuminant depend on the overall luminance of the checkerboard contexts. A higher estimated illuminant is returned for the high–high checkerboard context

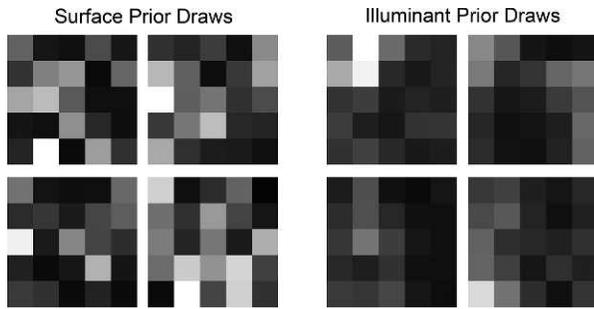


Figure 3. Example of equally likely draws from the derived surface prior (left) and the derived illuminant prior (right). The parameters for surfaces were $(\alpha_{surface}, \beta_{surface})$ and illuminant parameters were $(\mu_{illum}, \sigma_{illum}^2, \rho_{illum})$ as reported in the caption of Figure 2.

than for the standard checkerboard context, and similarly a higher estimated illuminant is returned for the standard checkerboard context than for the low–low checkerboard context. This also makes sense. Since surface reflectance is bounded in the prior to lie between 0 and 1, large changes in overall luminance must be caused by illumination changes.

Third, because the illumination estimates vary with checkerboard context, so too do the reflectance estimates. In particular, the reflectance estimate for the central test patch, which has the same luminance in each case, differs. Equally important, this estimate is affected by both the inner and outer ring luminances. This means that the algorithm predicts that both inner and outer ring luminance will affect perceived lightness.

To understand how the algorithm arrives at its estimates, it is helpful to consider the derived parameters of the prior distributions. Across possible sets of illuminants and surfaces consistent with a set of observed luminances, it is the priors that drive the algorithm estimates. Intuition about the derived priors is provided in pictorial form in Figure 3, which shows examples of draws from the derived prior distributions. Each checkerboard context represents a pattern of surfaces (left panels) or illuminants (right panels) of similarly high probability. Surfaces are independent from one another in the prior; thus, very dark surfaces are adjacent to very light surfaces in Figure 3 (left panels). The derived values of the beta distribution are such that darker surfaces are more probable than lighter surfaces. The prior distribution for the illuminant differs. Importantly, although the illuminant prior allows spatial variation, spatial variation that is gradual over locations is more likely than an abrupt change from one location to the next. The derived correlation parameter ρ_{illum} controls this aspect of the prior, with 1 indicating perfect correlation (uniform illumination) and 0 indicating independent illumination at each spatial location. The derived value of ρ_{illum} was 0.46, intermediate between these two extrema. In

Figure 3, the illuminants shown vary, but slowly, across space. The illuminant prior permits high and low illuminant luminance within one checkerboard context (top left illumination draw), but the highest illumination is not likely to be immediately adjacent to the darkest illumination. Relatively spatially uniform illuminations of different mean intensities are also probable (rightmost illumination draws).

Comparison of model predictions with human performance

Figure 4 shows a comparison of the CTFs predicted via the algorithm (lines) with those obtained from human observers (symbols). The panels are organized by the type of checkerboard manipulation, with inner ring manipulations in the top left, outer ring manipulations in top right, both ring manipulations of the same sign in the bottom left, and both ring manipulations of opposite sign in the bottom right. The solid black line in each panel represents the identity line: if checkerboard context had no effect on perceived lightness, the data (and algorithm estimates) would fall along this line.

Several salient characteristics of the model and psychophysical CTFs are clear from inspection of Figure 4. First, though neither the psychophysical nor model CTFs are straight lines in the log-log plots, they do have an average horizontal offset from the diagonal that varies with checkerboard context. For example, the luminance of the inner ring elicits a larger offset than the luminance of the outer ring (red and cyan lines further from the diagonal in the top left panel of Figure 4 than in the top right panel of Figure 4). The algorithm-based CTFs capture this inner–outer asymmetry. That regions of space close to a test influence its perception more than distant regions is a well-understood phenomenon, but algorithms that do not allow spatial variation in the illuminant cannot easily account for this phenomenon, since they are spatially stationary in terms of regional influence (but see Brainard et al., 2006 for an ad hoc approach).

Second, some CTFs exhibit an additional offset asymmetry: decreasing the luminance of the checkerboard context has a larger effect on perception than increasing the luminance (red lines further than cyan lines in the top right and bottom left panels of Figure 4). In contrast to the inner–outer asymmetry, it is not obvious from the luminance manipulations why this asymmetry should exist in some contexts but not in others (see Allred et al., 2012 for more discussion). The algorithm-based CTFs, however, do capture this broad feature of the data.

A third clear feature of the CTFs is that for high luminance test patches (right portions of each panel in

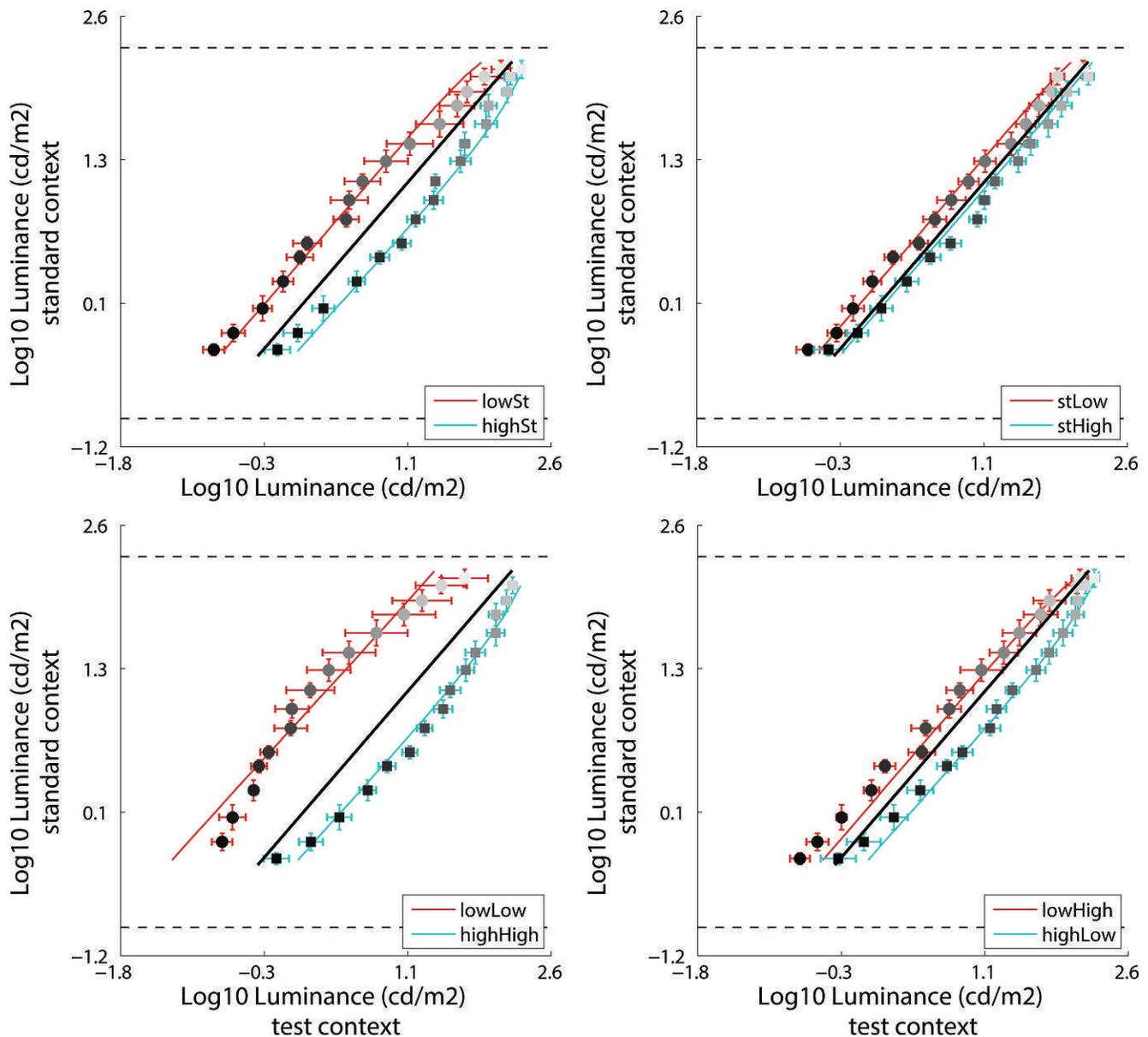


Figure 4. Psychophysical (data points) and algorithm-based (colored lines) CTFs. Each data point represents the average of the test patch values matched to a different Munsell paper in a test checkerboard context (x-axis) and the standard context (y-axis). The top left panel shows data for the low inner, standard outer (red) and high-inner, standard outer (cyan) test contexts; the top right panel shows data for the standard inner, low outer (red) and standard inner, high outer (cyan) contexts; the bottom left panel shows data for the low inner, low outer (red) and high-inner, high outer (cyan) test contexts; the bottom right panel shows data for the low inner, high outer (red) and high inner, low outer (cyan) contexts. Error bars are *SEM* across observers. Solid colored lines are CTFs computed using the Bayesian algorithm with derived surface prior parameters $\alpha_{surface} = 1$, $\beta_{surface} = 2$, and derived illuminant prior parameters $\mu_{illum} = 1$, $\sigma_{illum}^2 = 0.81$, $\rho_{illum} = 0.46$. The solid black identity line in each panel shows where the data would fall if there were no effect of context on lightness. The dashed horizontal lines represent the minimum and maximum test luminance.

Figure 4), there is a tendency of the data to curve toward the identity line. For checkerboard contexts with decreased luminance (red lines), this curvature indicates that larger changes in luminance in the test context are mapped to relatively smaller changes in luminance in the standard context. Another way to think about this phenomenon is that in these contexts, the perceived lightness (as indicated by the perceptually

equivalent standard context luminance) tends to saturate, so that a larger range of test patch luminances appears whitish. Just the opposite is true when checkerboard context luminance is increased. In these cases (cyan lines), small luminance changes in the test context are mapped to relatively larger luminance changes in the standard context. Again, another way to understand this is that in these contexts, a smaller range

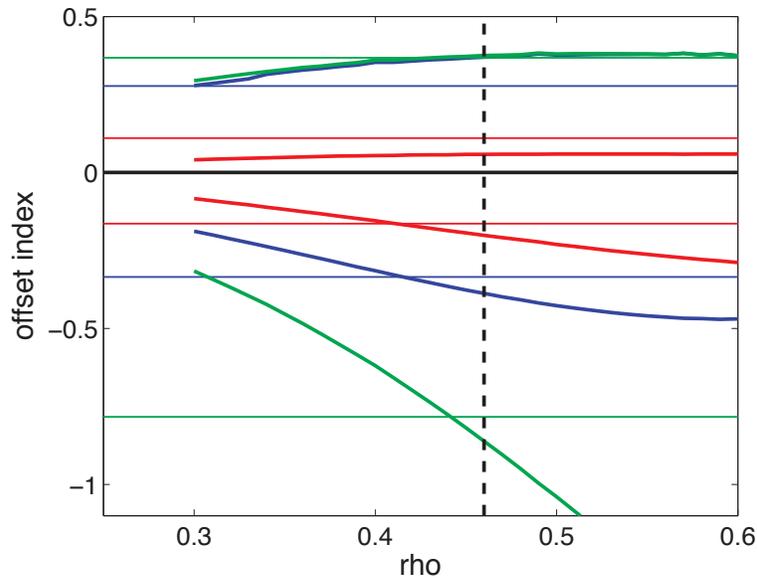


Figure 5. Effect of illuminant correlation on average CTF offset. The x-axis shows ρ_{illum} in the illuminant prior and the y-axis shows the offset index. The offset index is computed as the average horizontal distance (in log luminance) of psychophysical (thin lines) or model (heavy lines) CTFs from the diagonal. To obtain model CTFs from which to calculate the offset index, we computed algorithm estimates for different sets of prior parameters, where ρ_{illum} (illumination correlation parameter) took 31 values equally spaced between 0.3 and 0.6, and the other prior parameters were as reported above. CTFs were computed for checkerboard contexts in which luminance of the inner (blue), outer (red), or both (green) rings was increased (offsets above 0) or decreased (offsets below 0). Vertical black dashed line represents ρ_{illum} used to obtain the algorithm CTFs shown in Figure 4, and the horizontal black bar is 0 offset, for reference.

of test luminances tend to look whitish. The algorithm-based CTFs reproduce this curvature reasonably well for cases in which the contextual luminance tends to increase (cyan points and lines), although not for the cases in which the contextual luminance tends to decrease (red points and lines).

Finally, for some checkerboard contexts, there is also curvature at the lower end of the test patch range. This is most obvious in the low–low checkerboard contexts (red data in the bottom left panel). The algorithm-based CTFs fail to capture this curvature.

Understanding the average offset

The asymmetry of the horizontal offset in the CTFs caused by manipulating the inner and outer rings can be understood through the spatial variation in the illuminant. Because the algorithm is noise free, the reflectance estimates of the test patch (and hence the CTFs) are determined by the illuminant estimates at the test patch; thus, large changes in the test patch illuminant estimation between contexts result in large average offsets. Slow spatial variation in the algorithm's estimated illuminant means that estimates of test patch illumination will be coupled to illumination estimates of the entire checkerboard context. The spatial variation in the estimated illuminant is con-

trolled by the correlation parameter of the illuminant prior. Thus, to explore the algorithm's behavior with respect to the inner and outer rings, we manipulated the correlation parameter of the illumination prior and computed the effect of this manipulation on the average offset. To quantify this effect, we took as an offset index the average horizontal offset of each point of the CTF from the diagonal.

The strength of the correlation parameter in the illuminant prior has a systematic affect on the algorithm's offset index, as seen in Figure 5. We varied ρ_{illum} , holding other prior parameters constant, and computed algorithm-based CTFs for all checkerboard contexts. This offset index is shown in Figure 5 for the six checkerboard contexts in which the inner, outer, or both rings were increased or decreased. The solid horizontal lines represent the observed psychophysical offset in each condition. For each checkerboard context, more spatially correlated illuminant priors (higher ρ_{illum}) yield higher average offset values (all points tend away from 0 with increased ρ_{illum} in Figure 5). This effect is more pronounced when luminance is decreased relative to the standard checkerboard context (lines with offset indices less than 0). When ρ_{illum} is too low, the illuminant varies spatially too much and the context does not affect the test patch enough; that is, the algorithm offset from 0 is smaller than the average psychophysically measured offset.

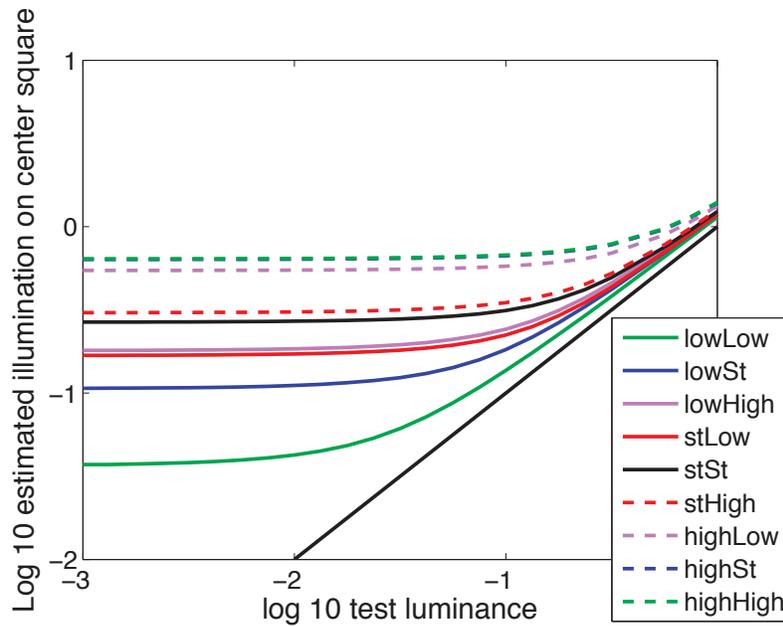


Figure 6. Effect of test patch luminance on overall illumination estimates. The y-axis shows the algorithm's estimated illumination for the center check as a function of test patch luminance for all nine checkerboard contexts; solid colored lines = low luminance profile checkerboard contexts; black line = standard context; dashed lines = high luminance profile checkerboard contexts. The thick black identity line shows the lower bound imposed on the illuminant estimate at the test location because of the constraint that surface reflectances do not exceed 1.

Understanding the curvature for high luminance tests

To understand the curvature for high luminance tests, it is helpful to consider how the algorithm's illuminant estimates are affected by the test patch luminance. Traditionally, tests in scenes are often thought of as probes that allow measurement of the overall effect of a particular context (Stiles, 1978), and the data are analyzed under the assumption that the presentation of the test does not itself play a contextual role. To the extent that this is true of the algorithm, estimates of the illuminant at the test location should be independent of test patch luminance. This in turn would make the algorithm-based CTFs lines with unit slope in the log-log plots shown in Figure 4. Instead, the algorithm-based CTFs exhibit some curvature, and this fact indicates that the algorithm's estimates of the illuminant at the test location within each context vary with test luminance and that this variation differs between contexts. To examine this further, Figure 6 plots the algorithm's estimate of the illuminant at the test patch location as a function of test patch luminance for each checkerboard context. For low test patch luminances, the assumption that the test itself does not play a contextual role holds true: the illumination estimate does not depend on test patch luminance. However, the test exerts a larger and larger effect as the test patch luminance increases (upward curvature in the right part of Figure 6). The size of the test-luminance effect varies with checkerboard context, and is more

pronounced for lower luminance checkerboards (solid lines).

It is intuitively clear why increasing the test luminance might increase the illuminant estimate at the test location: after all, when all else is equal, more light reflecting to the eye from a location is a likely indicator that more illumination was impinging on that location. Thus the upward curvature in Figure 6 is not surprising. Understanding what features of the algorithm create the detailed behavior of the curvature is harder, because the CTFs arise from a complex interaction between the surface prior, illuminant prior, and all 25 checkerboard luminances. One salient feature of the algorithm, however, is the constraint that surface reflectance estimate must be between 0 and 1. Since the likelihood is noise free, this constrains the illuminant estimates to be higher than the test luminance, shown by the thick black identity line in Figure 6. As test luminance increases, the estimated illumination in most contexts begins to approach this lower illumination bound and must curve upward to avoid crossing it. The upward curvature in Figure 6 in turn provides an explanation for the curvature in the predicted CTFs in Figure 4. Because the CTFs involve a comparison to the standard context, curvature in the CTFs will arise when the slopes of the colored lines in Figure 6 are different from the slope of the line representing the standard context. Thus we can understand the predicted CTF curvature seen at high

test luminances for the high–low and high–high contexts as a result of the fact that the illuminant lower bound does not affect these contexts as much as it does the standard context.

As we noted previously, the predicted CTFs do not capture the curvature seen in the psychophysical CTFs at low test luminances. With reference to Figure 6, this failure can be understood by noting that at low test luminances, the estimated illuminant is independent of test luminance for all checkerboard contexts. These illuminant estimates are not constrained by the bound shown by the thick black line and instead depend in a more complex way on the priors. Within our parametric model of surface and illuminant priors, we did not find parameters that could produce the appropriate predicted CTF curvatures at low test luminances while at the same time preserving an overall good fit to the psychophysical CTFs. It is possible that a different parametric choice of priors could remedy this aspect of our model's predictions.

Evaluating the overall model fit

The preceding plots and associated discussion indicate that the five-parameter Bayesian model captures much of the systematic variation in the data, but not all of it. To look at this quantitatively, we compared summary measures of the overall quality of the model fit to those obtained with a set of five comparison models. For each model, we summarized the overall quality of fit in two ways. The first was in terms of the overall root-mean-squared fit error to the entire data set of CTFs. This measure assesses how close the model predictions are, on aggregate, to the eight measured CTFs. The second was in terms of a root-mean-squared cross-validation error. The cross-validation error was obtained in a leave-one-out fashion, in which we fit the model to the data for each possible subset of six out of our seven observers, used the resulting parameters to predict the data for the left-out observer, and aggregated the prediction error over all seven left-out observers. The cross-validation measure is useful because it is sensitive to overfitting of the data by a model.

Three of our comparison models were not expected to provide a good fit to the data. Instead, these three models provided a sense for the variance in the data that were available to be modeled. The first model was an *overall mean* (OM) model, which fit the entire set of CTFs with their grand mean. The OM model errors represent an upper bound for any reasonable model. They also provide a sense of the total variance in the data set. In the second model, a *context mean* (CM) model, each CTF was fit by its own mean. The CM model errors provide a sense of the variance in the data

set that results from changing the test patch luminance, once the overall effect of checkerboard context has been modeled. Finally, in a *single-CTF* (SCTF) model, all eight CTFs were fit with the mean CTF. The SCTF model errors provide a sense of the variance in the data set that results from changing the checkerboard context, once the overall effect of test patch luminance has been modeled.

A fourth comparison model was a *nonparametric regression* (NP Reg) model, designed to provide an excellent description of the data (low overall fit error). The fits of this model were obtained using multivariate kernel smoothing regression (Nadaraya, 1964; Watson, 1964) with a Gaussian kernel, as implemented in the routine `ksrmv` made available by Yi Cao at the MATLAB Central File Exchange (<http://www.mathworks.com/matlabcentral/fileexchange>). This method, in essence, provides a smoothed look-up table of the data. We choose the width of the Gaussian kernel by hand to achieve a good overall fit. The overall fit error for NP Reg model is not of interest per se because it can be made very small by optimizing the parameters of the kernel regression. Further, such models do not provide any scientific insight about the nature of the computations mediating lightness perception. However, the cross-validation error for the NP Reg model is of interest because it provides a benchmark for other models. A competitor model that has a higher overall fit error than the NP Reg model can still in principle have a lower cross-validation error, depending on the degree to which the NP Reg model overfits the noisy data and the degree to which the competitor model captures structure in the data that survives measurement variability. Indeed, models that capture most of the underlying structure in the data should have cross-validation errors comparable to or lower than that of this model.

Our final comparison model was a *linear regression* (Lin Reg) model. This model predicts the CTFs as a linear function of the test and contextual log luminances. The model is a variant of the well-known retinex lightness algorithm (Land and McCann, 1971) and fit to our data. This follows because one of the standard variants of the retinex reduces to normalizing the test luminance by a spatially weighted geometric mean of all of the luminances in the image (Brainard & Wandell, 1986; Land, 1986). The Lin Reg model shares with our Bayesian model the fact that the CTFs are predicted directly from the image data, but with the Lin Reg model the predicted CTFs are constrained to be lines in the log-log plots of Figure 4. The Lin Reg model provides a reasonable benchmark for the performance of the Bayes model. Although we believe that the pursuit of Bayesian models of color and lightness is well-motivated theoretically (see Introduction and Discussion), it would reduce enthusiasm for

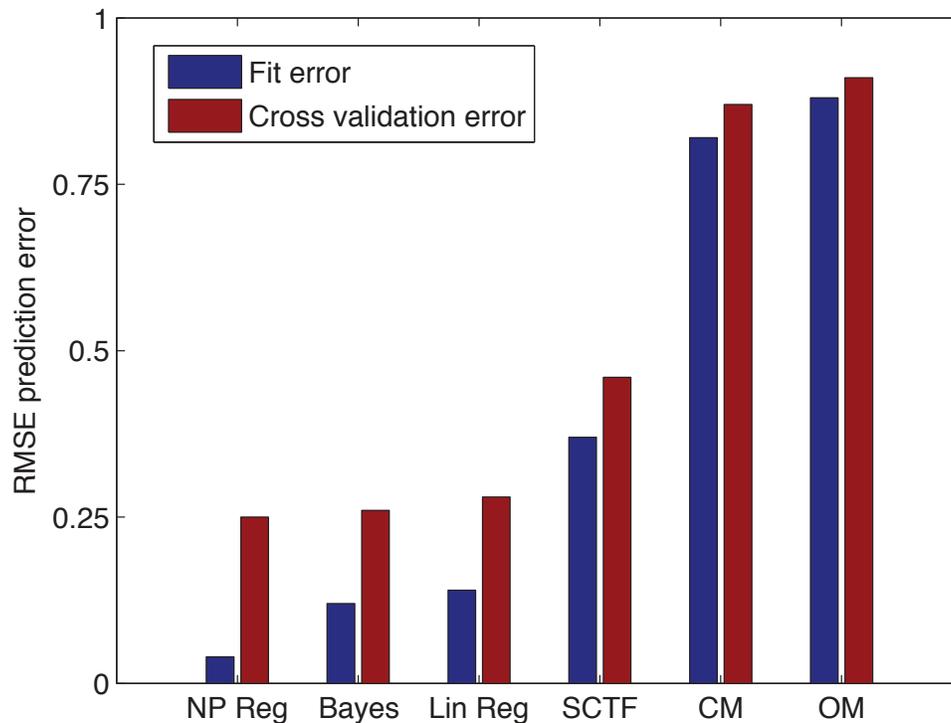


Figure 7. Overall fit errors (blue bars; see text for description) and cross-validation error (red bars; see text for description) for six different models: NP Reg (nonparametric regression), Bayes (Bayesian model), Lin Reg (linear regression), single-CTF (SCTF), CM (context mean), and OM (overall mean). The prediction error plotted is the root-mean-squared difference between the measured CTFs and those predicted by the relevant model. In fitting the Bayesian model to subsets of the data, we did not iterate over many possible starting points but rather began each search using the parameters that provided the best fit to the full data set.

further exploration if they cannot perform as well as extant more heuristically motivated models.

Figure 7 shows the overall fit error and cross-validation error for our Bayesian model (Bayes) and for the five comparison models. As expected, the OM, CM, and SCTF models all have high overall fit error and high cross-validation error. The NP Reg model has low overall fit error, but considerably higher cross-validation error. Of interest is that both the Bayes and Lin Reg models have cross-validation errors (0.26 and 0.28, respectively) similar to that of the NP Reg model (0.25). Given that both the Bayes and Lin Reg model produce smooth predicted CTFs that deviate from the measurements, it seems unlikely that these two models are overfitting the data. In addition, the similarity of their cross-validation errors with that of the NP Reg model suggests that the Bayes and Lin Reg models are capturing most of the overall variance in the data that survives individual differences.

The Bayesian model has slightly better overall fit and cross-validation error than the Lin Reg model (fit error 0.12 vs. 0.14; cross-validation error 0.26 vs. 0.28). This confirms, in an overall fit sense, the conclusions we drew above from examination of the data and Bayesian model fits Figure 4: the Bayesian model follows some of the curvature in the CTFs that is structurally incon-

sistent with the Lin Reg model. The differences between the two models in this regard are small, however.

Although the cross-validation analysis indicates that the Bayesian and linear regression models capture most of the overall variance that survives individual differences, a more fine-grained analysis of the cross-validation fits (not shown) reveals that the curvature shown in Figure 4 is a reliable feature of the data. This analysis examines the residuals of the predictions of each model as a function of the standard luminance, after shifting each CTF by the left-out subject so that the mean prediction matches the mean data. There is no obvious systematicity to the NP Reg model residuals when examined in this way, but the residuals for both the Bayes and Lin Reg models depend systematically on the standard luminance in a manner consistent with data shown in Figure 4. That is, once individual variability in the overall position of the CTFs is accounted for, the curvature seen in the aggregate data remains. The residual curvature is slightly smaller for the Bayesian model than for the Lin Reg model at high standard luminances, again consistent with the conclusions we drew above in our discussion of the full data set. The performance of the Bayesian model could thus be improved if it becomes possible to formulate priors that enable the model to better capture the

curvature seen in the data. In exploring the effect of the Bayesian models prior parameters on the predicted CTFs, we do find that there are sets of priors that yield curvature matching the data for some individual CTFs (e.g., the highly curved CTF measured for the low–low condition). These sets of prior parameters, however, yield very bad predictions for other CTFs.

Discussion

Spatial variation in the illuminant

The idea that the visual system is sensitive to spatial variation in the illuminant, and that the illuminant varies more slowly over space than do surfaces, was central to the retinex model of Land and McCann (1971). Unlike the current work, however, the computations driving the retinex model were based on heuristics and did not provide for explicit specification of image priors nor optimal use of image data for estimation. Work in computer vision also uses physical models in service of algorithms designed to separate illuminant and surface reflectance contributions for scenes that incorporate geometric structure (e.g., Barron & Malik, 2012; Bell & Freeman, 2001; Funt & Drew, 1988; Gehler, Rother, Kiefel, Zhang, & Scholkopf, 2011; Grosse, Johnson, Adelson, & Freeman, 2009; Romeiro & Zickler, 2010; Tappen, Freeman, & Adelson, 2005). Our Bayesian algorithm shares with this work the fact that the illumination is allowed to vary spatially. In our work, however, we exploited the structure of the restricted class of scenes we studied to simplify the formulation. We linked algorithm output to human performance and found that we can account for effects of overall changes in image luminance (Figure 4), the dependence of such changes on spatial location (Figure 4 and Figure 5), the curvature of the measured context transfer functions at high test luminances for some but not all CTFs (Figure 4 and Figure 6), and the observed increment–decrement asymmetries (Figure 4 and Figure 5). Although there are aspects of the data that our model does not account for, such as the curvature of the CTFs at low test luminances, the fact that it provides a unified account of much of the data suggests the usefulness of this approach for developing quantitative models of perceived lightness for spatially complex images.

To verify that our model's success at accounting for spatial effects was not overly specific to our stimuli, we applied it to simulations of the staircase Gelb effect, an illusion that has been taken as a challenge for models based on inverse-optics algorithms (Cataliotti & Gilchrist, 1995; Gilchrist, 2006). To produce the illusion, a series of papers that range from black to white is

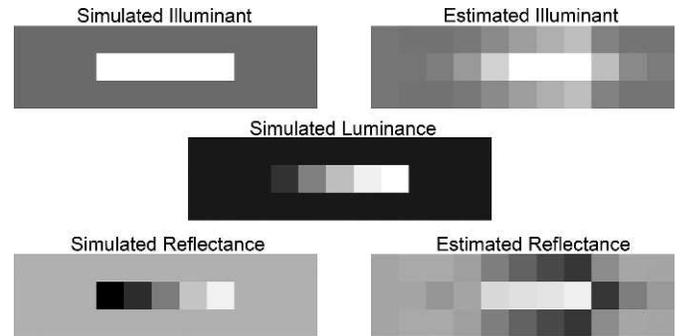


Figure 8. Algorithm estimates for the staircase Gelb effect. Leftmost panels include the simulated spotlight (top) and simulated surfaces ranging from black to white (bottom). Central panel is the simulated luminance, created by pixel-wise multiplication of the illuminant and reflectance. Values are from the staircase Gelb effect reported in (Cataliotti & Gilchrist, 1995). Papers span a 30:1 reflectance range (0.03 to 0.90), and the five-square spotlight is 30 times brighter than its surround. Right panels show algorithm estimates for the illuminant (top) and reflectance (bottom). Values have been scaled for visualization purposes. One scaling procedure is used for illuminants (both simulated and estimated, top panels), one for reflectance (both simulated and estimated, bottom panels), and one for luminance (center panel).

illuminated by a spotlight in an otherwise dim room. Though the papers range from black to white (reflectance contrast = 30:1) observers typically report that the papers appear to range from midgray to white (reflectance contrast = 3:1; Cataliotti & Gilchrist, 1995). This illusion cannot be explained by the idea that observers make an overall misestimation of a spatially uniform illuminant. Using the published values (Cataliotti & Gilchrist, 1995), we simulated the luminance of the physical setup and, using the Bayesian algorithm with the derived prior parameters reported here, estimated the reflectance and illumination at each location.

The algorithm's estimates at locations corresponding to staircase Gelb effect stimuli are in good agreement with the phenomenology of the illusion, as shown in Figure 8, in which the algorithm estimates the contrast ratio between the darkest simulated paper and the lightest simulated paper to be 1.97:1, although the actual luminance contrast under the simulated uniform spotlight is 30:1. To see why, it is helpful to consider the illumination estimate (Figure 8, top right panel). In the simulation, the luminance of the white paper under the spotlight is very high. Since reflectance of surfaces is constrained in the algorithm and in the real world, the algorithm must solve the ambiguity by estimating a very high illuminant. However, outside of the spotlight, the very low luminance values constrain the algorithm to estimate both low reflectance and low illumination. The illumination cannot change abruptly at the edge of the simulated spotlight, since the illumination prior is

constrained to vary slowly over space. Thus, the algorithm estimates that the illumination over the simulated papers is changing slowly and the darkest paper is under a lower illumination than the lightest paper. To be consistent with the luminance data, the algorithm must overestimate the reflectance of the darker papers, resulting in the observed compression.

We do note that the algorithm's estimates at locations outside of the five central squares do not correspond with the usual description of this illusion. For example, to the right of the most luminous square, the algorithm estimates a high illuminant and a correspondingly low surface reflectance. In the original paper, Cataliotti and Gilchrist (1995) did not measure the perceived lightness of surfaces immediately adjacent to the spotlight-illuminated papers. Indeed, in the actual illusion, a series of five papers is presented in isolation, with the surrounding surfaces at some depth behind the papers. However, though perceived lightness at the background locations was not measured, an abrupt darkening outside the spotlight is not a salient perceptual feature of the illusion. Thus, the algorithm's estimates probably do not predict human perception of the background. One reason for this may be the fact that we employed priors that enforce smoothly varying illumination and do not model the possibility of sharp illumination boundaries. This is a limitation that we suspect could be overcome by employing illuminant priors that enforce piece-wise rather than global smoothness of the illumination, and exploring the effect of incorporating such priors is of interest for future work. Methods for specifying and computing with these general types of priors are available (Geman & Geman, 1984; Kersten, 1991; Li, 2001; Simoncelli, 2005). A second elaboration would be to relax the assumption that the surface reflectances at each checkerboard location are independent.

Across our contexts, we only manipulated the luminance of the checkerboard contexts surrounding the test. It is clear that geometric information also plays a role in how the visual system segments an image into different regions of illumination (Adelson, 1993, 2000; Bloj et al., 2004; Boyaci, Maloney, & Hersh, 2003; Gilchrist, 1980, 2006; Gilchrist et al., 1999; Hochberg & Beck, 1954; Ripamonti et al., 2004; see also Lee & Smithson, 2012). This includes cues to the three-dimensional structure of the scene that might provide information about how the illumination varies across image locations as well two-dimensional image features that might indicate illumination boundaries (e.g., the spatial structure of junctions identified in the image). Our priors do not model the three-dimensional structure of objects nor the three-dimensional geometry of illumination, nor does our likelihood describe the relation between three-dimensional scenes and two-dimensional images. Thus our algorithm is not sensitive

to cues about the three-dimensional structure of the scene nor to geometric structure in the image other than the distance between locations. For this reason, it is clear a priori that our model will not account for the type of geometric effects on lightness described in the references listed above. Currently, our quantitative understanding of such perceptual effects is still in its relative infancy, as is our understanding of how photometric and geometric information interact as the visual system segments the image (but see Lee & Brainard, 2011 for some initial work on the latter question). As our understanding and ability to compute evolves, it should be possible to develop Bayesian models of lightness that incorporate additional geometric factors (see for example Barron & Malik, 2012; Romeiro & Zickler, 2010).

Test effect

A simplifying assumption often made in studying the effect of context on lightness is that the test itself does not substantially perturb the context. When correct, this means that data collected across different test luminances may be interpreted as characterizing a single fixed context (Stiles, 1978). This assumption, however, may not be secure. As discussed above, the curvature in the psychophysical CTFs may be understood as a perturbation of the context by the test itself. Thus the test, rather than neutrally probing the effect of the surrounding context on the visual system's processing of light at the test location, actually alters the visual system's state at that location. We think it is important for theorists to keep this possibility in mind, and note that this type of effect can be incorporated, albeit imperfectly to date, into the type of Bayesian model we develop here.

Relation to other work

The modeling approach we have taken here is part of a broader program that aims to relate visual performance to the solution of estimation problems that the visual system must solve to convert ambiguous sense data into useful perceptual representations (Brainard, 2009; Brainard & Maloney, 2011; Geisler, 2011; Kersten, Mamassian, & Yuille, 2004; Knill & Richards, 1996; Morgenstern, Murray, & Harris, 2011; Purves & Lotto, 2003; Rust & Stocker, 2010; Stocker & Simoncelli, 2006; Weiss, Simoncelli, & Adelson, 2002). Although we believe this is a useful way to approach understanding lightness perception, it is not the only approach. A complementary effort attempts to relate perceived lightness to the action of psychophysical mechanisms that abstract key features of the underly-

ing physiology of the visual pathways. Examples of this approach in the lightness domain include work by Blakeslee and McCourt (1999), Blakeslee and McCourt (2004), Chubb, Sperling, and Solomon (1989), Radonjić et al. (2011), and Rudd and Zemach (2004).

Indeed, in our earlier report (Allred et al., 2012) we showed that the parametric variation in the shapes of the CTFs that we consider here is very well accounted for by a model that connects lightness to the response of a saturating visual mechanism (Radonjić et al., 2011). What that model did not provide is an account of how to derive the response function parameters for any context from a description of the patch luminances in that context. Thus, it is not possible to use that model to predict lightness for contexts beyond those that were directly measured. The current work, in contrast, focuses on how and why any particular context exerts its influence on the CTFs. An interesting direction for future research may be to try to incorporate known features of visual physiology, such as the fact that mechanisms have limited dynamic range, into the formulation of otherwise optimal estimation algorithms. Additional discussion of the relation between computational and mechanistic approaches to understanding constancy is available elsewhere (Brainard, 2004, 2009; Foster, 2011; Maloney & Brainard, 2010; Pokorný, Shevell, & Smith, 1991; Smithson, 2005).

Concluding remarks

The information available to the visual system through the retinal image is ambiguous; any functional understanding of visual perception must account for how this ambiguity is resolved. Lightness perception provides a model system for studying how the visual system resolves ambiguity. Here we show that an illuminant-surface estimation method that combines image luminances with priors that capture environmental statistical regularities can account for many of the broad features of a large empirical data set on lightness perception.

To evaluate our algorithm-based model, we fit algorithm-based CTFs to CTFs obtained from judgments of the perceived lightness of test patches embedded in grayscale checkerboard contexts (Allred et al., 2012). These checkerboard stimuli incorporated the large variation in luminance that is a pervasive feature of natural scenes (Heckaman & Fairchild, 2009; Mury, Pont, & Koenderink, 2009; Xiao, DiCarlo, Catrysse, & Wandell, 2002). In addition, the luminance profile of the checks both near to and remote from the central test patches was systematically manipulated. The manipulations provided a simplified version of the kind of spatial changes in illumination that occur in

real scenes. The algorithm-based model accounts for the broad features of the data and some but not all of the more detailed features of the data.

The performance of our algorithm is driven primarily by priors over illumination and surfaces that it incorporates. We choose simple parametric forms for these priors, with the form of the illuminant being a multivariate log normal that allowed expression of the intuition that illumination varies more slowly over space than surface reflectance. This is a reasonable point of departure. More sophisticated priors would incorporate spatial structure for both surfaces and illuminants and characterize the nature of the spatial variation for both in more detail than can be described by the first-order correlation structure alone (Geman & Geman, 1984; Kersten, 1991; Li, 2001; Simoncelli, 2005). Formulating the priors over a three-dimensional representation of the scene and allowing the likelihood to map between this representation and the image data is a related and important direction for future research (see, for example, Barron & Malik, 2012; Romeiro & Zickler, 2010).

The prior parameters used to model the data were obtained as those that provided the best model fit to the psychophysical data. We refer to these as the derived priors. The derived priors can be understood as those that are brought to bear by the visual system, within the context of our model and experimental stimuli (cf. Brainard et al., 2006; Stocker & Simoncelli, 2006; Brainard, Williams, & Hofer, 2008; Morgenstern et al., 2011; Girshick, Landy, & Simoncelli, 2011). As such, they provide an interpretable description of human performance, again within the context of our model. In particular, the derived priors characterize human performance in the currency of the statistical structure of natural scenes, and as such it would be interesting to know how closely the derived priors match the priors obtained directly from physical measurements of natural scenes (see Girshick et al., 2011 for such comparison in the perceptual domain of spatial orientation and Allred, 2012 for a general discussion). We are currently limited, however, in terms of what we know about the relevant natural scene statistics. Although there are several valuable data sets of calibrated natural images now available, these image data sets do not allow separate characterization illuminant and surface reflectance statistics (e.g., Chakrabarti & Zickler, 2011; Foster, Nascimento, & Amano, 2004; Heckaman & Fairchild, 2009; Mury et al., 2009; Olmos & Kingdom, 2004; Parraga, Brelstaff, Troscianko, & Moorehead, 1998; Tkacik et al., 2011; van Hateren & van der Schaaf, 1998; Xiao et al., 2002). Similarly there is work on the geometrical structure of natural illumination fields (Debevec, 1998; Dror, Will-sky, & Adelson, 2004; Morgenstern et al., 2011), but translating the characterization provided by this work to image plane statistics is nontrivial. As we obtain

better measurements of the distribution of surface reflectances and illumination intensities in natural scenes, it may become possible to both improve upon our choice of prior parametric forms and to make informative comparisons between priors derived from analysis of human performance and their counterparts obtained directly from physical measurements.

Keywords: Bayesian, lightness/brightness perception, luminance, high dynamic range display, illumination

Acknowledgments

Supported by NIH RO1 EY10016 and NIH P30 EY001583 to David Brainard and NSF BCS 0954749 to Sarah Allred.

Commercial relationships: none.

Corresponding author: Sarah R Allred.

Email: srallred@camden.rutgers.edu.

Address: Department of Psychology, Rutgers, The State University of New Jersey, Camden, NJ, USA.

References

- Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press.
- Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, *3*, 2042–2044.
- Allred, S. R. (2012). Approaching color with Bayesian algorithms. In G. Hatfield & S. R. Allred (Eds.), *Visual experience: Sensation, cognition, and constancy* (1st ed., chap. 11). Oxford: Oxford University Press.
- Allred, S. R., Radonjić, A., Gilchrist, A. L., & Brainard, D. H. (2012). Lightness perception in high dynamic range images: Local and remote luminance effects. *Journal of Vision*, *12*(2):7, 1–16, <http://www.journalofvision.org/content/12/2/7>, doi:10.1167/12.2.7. [PubMed] [Article]
- Barron, J., & Malik, J. (2012). *Color constancy, intrinsic images, and shape estimation*. Firenze, Italy: European Conference on Computer Vision.
- Bell, M., & Freeman, E. (2001). Learning local evidence for shading and reflectance. *Computer Vision, 2001. ICCV 2001*, *1*, 670–677.
- Berger, T. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.
- Blakeslee, B., & McCourt, M. E. (1999). A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction. *Vision Research*, *39*(26), 4361–4377.
- Blakeslee, B., & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Research*, *44*(21), 2483–2503.
- Bloj, M., Ripamonti, C., Mitha, K., Greenwald, S., Hauck, R., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal of Vision*, *4*(9):6, 735–746, <http://www.journalofvision.org/content/4/9/6>, doi:10.1167/4.9.6. [PubMed] [Article]
- Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision*, *3*(8):2, 541–553, <http://www.journalofvision.org/content/3/8/2>, doi:10.1167/3.8.2. [PubMed] [Article]
- Brainard, D. H. (2009). Bayesian approaches to color vision. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 395–408). Cambridge, MA: MIT Press.
- Brainard, D. H. (2004). Color constancy. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 948–961). Cambridge, MA: MIT Press.
- Brainard, D. H., Kraft, J. M., & Longère, P. (2003). Color constancy: Developing empirical tests of computational models. In *Colour perception: Mind and the physical world* (pp. 307–334). Oxford: Oxford University Press.
- Brainard, D. H., Longere, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., & Xiao, B. (2006). Bayesian model of human color constancy. *Journal of Vision*, *6*(11):10, 1267–1281, <http://www.journalofvision.org/content/6/11/10>, doi:10.1167/6.11.10 [PubMed] [Article]
- Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illuminant models. *Journal of Vision*, *11*(5):1, 1–18, <http://www.journalofvision.org/content/11/5/1>, doi:10.1167/11.5.1. [PubMed] [Article]
- Brainard, D. H., & Radonjić, A. (in press). Color constancy. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (2nd ed.). Cambridge, MA: MIT Press.
- Brainard, D. H., & Wandell, B. A. (1986). Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, *3*, 1651–1661.
- Brainard, D. H., Williams, D. R., & Hofer, H. (2008). Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color

- appearance of small spots. *Journal of Vision*, 8(5): 15, 1–23, <http://www.journalofvision.org/content/8/5/15>, doi:10.1167/8.5.15. [PubMed] [Article]
- Cataliotti, J., & Gilchrist, A. (1995). Local and global processes in surface lightness perception. *Perception and Psychophysics*, 57(2), 125–135.
- Chakrabarti, A., & Zickler, T. (2011). Statistics of real-world hyperspectral images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 193–200). Providence, RI: Conference proceedings.
- Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences, USA*, 86(23), 9631–9635.
- Debevec, P. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Siggraph* (pp. 189–198). New York: ACM.
- Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004). Statistical characterization of real-world illumination. *Journal of Vision*, 4(9):11, 821–837, <http://www.journalofvision.org/content/4/9/11>, doi:10.1167/4.9.11. [PubMed] [Article]
- Foster, D. H. (2011). Color constancy. *Vision Research*, 51, 674–700.
- Foster, D. H., Nascimento, S. M. C., & Amano, K. (2004). Information limits on neural identification of colored surfaces in natural scenes. *Visual Neuroscience*, 21, 1–6.
- Funt, B. V., & Drew, M. S. (1988). Color constancy computation in near-mondrian scenes using a finite dimensional linear model. In *Proceedings CVPR '88* (pp. 544–549). Simon Fraser University, Burnaby, BC: Computer Society Conference.
- Gehler, P., Rother, C., Kiefel, M., Zhang, L., & Scholkopf, B. (2011). Recovering intrinsic images with a global sparsity prior on reflectance. In J. Shawe-Taylor, R. S. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger, *Advances in neural information processing systems 24*. NIPS: <http://books.nips.cc/nips24.html>.
- Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, 51, 771–781.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Gilchrist, A. (2006). *Seeing black and white*. Oxford: Oxford University Press.
- Gilchrist, A. (1980). When does perceived lightness depend on perceived spatial arrangement? *Perception and Psychophysics*, 28, 527–538.
- Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., et al. (1999). An anchoring theory of lightness perception. *Psychological Review*, 106(4), 795–834.
- Girshick, A., Landy, M., & Simoncelli, E. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, 14(7), 926–932.
- Grosse, R., Johnson, M. K., Adelson, E. H., & Freeman, W. T. (2009). Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *Computer Vision, 2009 IEEE 12th International Conference*, Kyoto, Japan (pp. 2335–2342).
- Heckaman, R. L., & Fairchild, M. D. (2009). Jones and condit redux in high dynamic range and color. In *Seventeenth Color Imaging Conference: Color Science and Engineering Systems, Technologies and Applications* (pp. 8–14). Albuquerque, NM: Society for Imaging Science and Technology.
- Helmholtz, H. (1910). *Helmholtz's physiological optics*. New York: Optical Society of America. (Original work published 1867).
- Hochberg, J. E., & Beck, J. (1954). Apparent spatial arrangement and perceived brightness. *Journal of Experimental Psychology*, 47, 263–266.
- Kersten, D. (1991). Transparency and the cooperative computation of scene attributes. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 209–228). Cambridge, MA: MIT Press.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as bayesian inference. *Annual Review of Psychology*, 55, 271–304.
- Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.
- Land, E. H. (1986). Recent advances in retinex theory. *Vision Research*, 26, 7–21.
- Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1), 1–11.
- Lee, P. M. (1989). *Bayesian statistics*. London: Oxford University Press.
- Lee, R. J., & Smithson, H. E. (2012). Context-dependent judgments of color that might allow color constancy in scenes with multiple regions of illumination. *Journal of the Optical Society of America (A)*, 29, A247–A257.
- Lee, T. Y. & Brainard, D. H. (2011). Detection of

- changes in luminance distributions. *Journal of Vision*, 10(13):14, 1–16, <http://www.journalofvision.org/content/11/13/14>, doi:10.1167/11.13.14. [PubMed] [Article]
- Li, S. Z. (2001). *Markov random field modeling in image analysis*. Tokyo: Springer-Verlag.
- Maloney, L. T., & Brainard, D. H. (2010). Color and material perception: achievements and challenges. *Journal of Vision*, 10(9):19, 1–6, <http://www.journalofvision.org/content/10/9/19>, doi:10.1167/10.9.19. [PubMed] [Article]
- Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. *Proceedings of the National Academy of Science, USA*, 108(30), 12551–12553.
- Mury, A. A., Pont, S. C., & Koenderink, J. J. (2009). Structure of light fields in natural scenes. *Applied Optics*, 48(28), 5386–5395.
- Nadaraya, E. (1964). On estimating regression. *Theory of Probability and its Applications*, 9(1), 141–142.
- Olmos, A., & Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, 33, 1463–1473.
- Parraga, C. A., Brelstaff, G., Troscianko, T., & Moorehead, I. R. (1998). Color and luminance information in natural scenes. *Journal of The Optical Society of America A*, 15, 563–569.
- Pokorny, J., Shevell, S., & Smith, V. (1991). Colour appearance and colour constancy. In P. Gouras (Ed.), *Vision and visual dysfunction: Vol. 6. The perception of colour* (pp. 43–61). London: Macmillan.
- Purves, D., & Lotto, R. B. (2003). *Why we see what we do: An empirical theory of vision*. Sunderland, MA: Sinauer.
- Radonjić, A., Allred, S. R., Gilchrist, A. L., & Brainard, D. H. (2011). The dynamic range of human lightness perception. *Current Biology*, 21(22), 1931–1936.
- Ripamonti, C., Bloj, M., Hauck, R., Kiran, M., Greenwald, S., Maloney, S. I., et al. (2004). Measurements of the effect of surface slant on perceived lightness. *Journal of Vision*, 4(9):7, 747–763, <http://www.journalofvision.org/content/4/9/7>, doi:10.1167/4.9.7. [PubMed] [Article]
- Romeiro, F., & Zickler, T. (2010). *Inferring reflectance under real-world illumination* (Technical Report No. TR-10-10). Cambridge, MA: Harvard School of Engineering and Applied Sciences.
- Rudd, M. E., & Zemach, I. K. (2004). Quantitative properties of achromatic color induction: an edge integration analysis. *Vision Research*, 44(10), 971–981.
- Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: two fundamental challenges for visual processing. *Current Opinion in Neurobiology*, 20(3), 382–388.
- Shevell, S. K., & Kingdom, F. A. A. (2008). Color in complex scenes. *Annual Review of Psychology*, 59, 143–166.
- Simoncelli, E. P. (2005). Statistical modeling of photographic images. In A. Bovik (Ed.), *Handbook of image and video processing* (pp. 431–441). Salt Lake City, UT: Academic Press.
- Smithson, H. E. (2005). Sensory, computational, and cognitive components of human color constancy. *Philosophical Transactions of the Royal Society of London B*, 360, 1329–1346.
- Stiles, W. S. (1978). Introductory essay. Increment thresholds in the analysis of colour-sensitive mechanisms of vision: Historical retrospect and comment on recent developments. In *Mechanisms of colour vision* (pp. 1–34). London: Academic Press.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4), 578–585.
- Tappen, M., Freeman, W., & Adelson, E. (2005). Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9), 1459–1472.
- Teller, D. Y. (1984). Linking propositions. *Vision Research*, 24(10), 1233–1246.
- Tkacik, G., Garrigan, P., Ratliff, C., Milcinski, G., Klein, J. M., Sterling, P., et al. (2011). Natural images from the birthplace of the human eye. *PLoS ONE*, 6(6), e20409.
- van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings: Biological Sciences*, 265(1394), 359–366.
- Watson, G. (1964). Smooth regression analysis. *The Indian Journal of Statistics, Series A*, 26(4), 359–372.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6), 598–604.
- Xiao, F., DiCarlo, J., Catrysse, P., & Wandell, B. (2002). *High dynamic range imaging of natural scenes*. Scottsdale, AZ: Conference proceedings from the 10th Color Imaging Conference: Color Science, Systems and Applications.