Modeling lightness and color perception of high dynamic range stimuli

Michael E. Rudd; University of Nevada, Reno, NV USA

Abstract

A model of lightness computation by the human visual system is described and simulated. The model accounts to within a few percent error for the large perceptual dynamic range compression observed in lightness matching experiments conducted with Staircase Gelb and related stimuli [1]. The model assumes that neural lightness computation is based on transient activations of ON- and OFF-center neurons in the early visual pathway during the course of fixational eye movements. The receptive fields of the ON and OFF cells are modeled as difference-of-gaussian functions operating on a log-transformed image of the stimulus produced by the photoreceptor array. The key neural mechanism that accounts for the observed degree of dynamic range compression is a difference in the neural gains associated with ON and OFF cell responses. The ON cell gain is only about ¼ as large as that of the OFF cells. ON and OFF cell responses are sorted in visual cortex by the direction of the eye movements that generated them, then summed across space by large-scale receptive fields to produce separate ON and OFF edge induction maps. Lightness is computed by subtracting the OFF network response at each spatial location from the ON network response and normalizing the spatial lightness representation such that the maximum activation within the lightness network always equals a fixed value that corresponds to the white point. In addition to accounting for the degree of dynamic range compression observed in the Staircase Gelb illusion, the model also accounts for change in the degree of perceptual compression that occurs when the spatial ordering of the papers is altered, and the release from compression that occurs when the papers are surrounded by a white border. Furthermore, the model explains the Chevreul illusion [2] and perceptual fading of stabilized images [3,4] as a byproduct of the neural lightness computations assumed by the model.

The Staircase Gelb Stimulus: An HDR Display

The difference between the ambient light level on a bright summer day to that of a just barely detectable stimulus presented to the fully dark-adapted eye has been estimated to be at least 9 log units [5]. Many different neural mechanisms work in tandem to map this enormous range of light levels to the much smaller range of physiological response magnitudes. These include shifts between rod and cone-mediated vision, luminance- and contrast- based adaptation, and lightness anchoring. Here, I describe the results of computer simulations of a neural model that accounts with high precision for the perceptual dynamic range compression that is observed when participants make judgments of the lightness (perceived reflectance) of achromatic surfaces illuminated by a spotlight in an otherwise dimly lit room. The model accounts for the magnitude of this compression by assuming a simple gain control mechanism through with the neural gains applied to local luminance increments and decrements differ from one another, such that the gain applied to increments being only about ¼ as large as the gain applied to decrements. Surface lightness is computed by integrating these gain-controlled neural responses to incremental and decremental luminance across space with large-scale receptive fields in visual cortex.

The proposed mechanism by which the ON and OFF cell responses are generated depends on fixational eye movements: the small, random, eye movements that occur when the observer’s eyes are “fixated” on an image. During the course of these eye movements, ON-center and OFF-center cells in the early visual pathway are transiently activated when their receptive fields translate across luminance edges in the stimulus. The magnitudes of these activations are “read out” within a sensory integration window of 100 msec or less and sorted in visual cortex to generate separate cortical maps of the activations corresponding to each eye movement direction. At a subsequent stage of cortical processing, the ON and OFF activations are independently integrated across space by large-scale receptive fields in a way that takes into account the degree to which the eye movement that generated each ON or OFF activation pointed in the direction of the large-scale receptive field center. This spatial integration results in two maps representing ON and OFF spatial induction signals. At the final stage of the model, activities these maps are linearly combined to produce a single neural map of perceive reflectance. This map is normalized so that the highest activation in the map always equals a constant value corresponding to the percept of white.

The process by which the ON and OFF induction maps are combined depends on the output of a high-level pointwise leaky temporal integration of the maps computed within each sensory integration period. One implication of this leaky integration is that the neural image of perceived reflectance that is generated at the output stage of the model will fade perceptually if the eye movements are artificially halted, as has been demonstrated to occur in several psychophysical studies [3,4]. The time that it takes for this perceptual fading to occur corresponds, in the model, to the characteristic time of the leaky integrator that temporally integrates the neural activities within the ON and OFF induction maps across multiple sensory integration periods. This characteristic time is estimated from psychophysical data to be a few seconds in duration.

In what follows, I first describe the high dynamic range displays and lightness matching procedures that were used to test the neural model against perceptual data, then I describe the neural lightness computation model in more detail. Finally, I present the result of computer simulations of the model’s response to the HDR displays, and I demonstrate that the model also generates the Chevreul illusion, in which a series of bands of homogeneous luminance, ordered from lowest to highest luminance, takes on a scalloped appearance in brightness.

Simulated Lightness Experiments

In the Staircase Gelb lightness paradigm, a series of achromatic surfaces (e.g. papers) are arranged in order from darkest to lightest in a spotlight, and an observer judges the lightness of each paper with a matching technique. Two noteworthy perceptual
phenomena are observed in this experiment. First, the paper with the high reflectance (luminance) always appears white, regardless of the paper’s actual physical reflectance [6]. When a Munsell matching technique is used, the highest luminance surface is typically matched to a Munsell 5.0 standard. Second, the range of the perceived surface reflectances across the series is highly compressed relative to the range of physical reflectances. For example, the range of judged reflectances of the papers in the original Staircase Gelb experiment of Cataliotti and Gilchrist [7] was only about 1/3 as large, in log units, as the range of the actual reflectances [8]. A replication of Gilchrist and Cataliotti’s by Zavagno, Annan, and Caputo (Fig. 1) produced a nearly identical degree of dynamic range compression [1].

Cataliotti and Gilchrist proposed an explanation of their results based on lightness anchoring theory, which proposes that the lightness of each surface in the series is determined by the ratio of that paper’s luminance relative to that of the highest luminance paper. Zavagno et al. tested this theory by replicating the Staircase Gelb experiment (their Series A in Fig. 1), then repeated the experiment with some of the papers spatially reordered. In Series B, the highest luminance paper was position in between the lowest luminance paper and the second to the lowest luminance paper. In Series C, the highest luminance paper was repositioned on the opposite end of the series, next to the lowest luminance paper. The purpose of these reorderings was to investigate whether appearance of the papers would be influenced by contrast with respect to the neighboring papers, in addition to by the paper’s relationship to the highest luminance paper.

The papers in the Zavagno et al. study were each 2.38 deg x 2.38 deg. They were illuminated by a theatrical spot lamp positioned behind the observer in an otherwise dimly lit room. The observer judged the lightness of each paper by choosing the Munsell chip that best matched the paper’s apparent reflectance.

Fig. 1 shows that the spatial arrangement of the papers did influence their lightness. The first goal of the simulations present here was to demonstrate that the lightness computation model described above can account quantitatively for these results.

In their original experiments with the Staircase Gelb display, Gilchrist and Cataliotti [9] further examined the effects of surrounding the entire series of papers with a white border. This was found to substantially relieve the perceptual compression that occurred when the papers were presented in isolation against a dark background, resulting in a more veridical perceptual scaling (i.e., more in line with the dashed “ground truth” line in Fig. 1). This effect was also simulated in the present study.

Details of the Simulation Model

Input Array

The input array was 3610 x 5210 pixels (rows by columns). Centered within this array was a representation of the five-paper Staircase Gelb series (Series A). Each “paper” in the series was 477 x 477 pixels, which corresponded to a scaling factor of 200 pixels/deg with one extra pixel added per side of each paper so that a readout at numerical readout at the output layer of the model could be based on a single pixel located in the paper’s center. The paper luminances were 3.12, 12.0, 30.0, 59.1, and 90.0 cd/m², and the luminance of the background field was 2.82 cd/m².

Random Walk Model of Fixational Eye Movements

A standard model of fixational eye movements is Brownian motion in two dimensions [11]-[13], which was approximated here by a discrete 2D random walk. Since eye movements function in the lightness model to detect edges as they traverse edges within the input image, and the papers in the display were squares, it sufficed to model only four types of eye movements: left, right, up, and down. The direction of each eye movement was chosen randomly and independently of the direction on any other time step. Eye movements occurred once every 30 time steps, except when the eye movements were halted in the simulation to model the perceptual fading of stabilized images.

Photoreceptor Array

A photoreceptor array encoded the luminance of each pixel within the input array once per time step. The temporal response of the photoreceptors was assumed to be instantaneous. The magnitude of the photoreceptor response to each pixel was computed as the logarithm of the input luminance at that pixel’s location. Thus, the magnitudes of the photoreceptor output within the five squares were 0.4942, 1.0792, 1.4778, 1.7716, and 1.9543; and within the background it was 0.49.

As a consequence of the eye movements, the locations of the

![Figure 1. Dynamic range compression observed in the lightness matching experiment of Zavagno, Annan, and Caputo. (See text for explanation).](image1)

![Figure 2. Example of the photoreceptor array output on one time step of the simulation.](image2)
five squares within the photoreceptor array underwent a random walk, with a displacement of 12 pixels occurring in a random direction (up, down, left, or right) once every 30 time steps. An example of one temporal frame of the photoreceptor array output is shown in Fig. 2. The darkest paper is somewhat difficult to see here because its luminance in log units is close to that of the background.

**Transient ON and OFF Cell Responses**

At the next stage of the model, the random translations of the photoreceptor array produce transient ON and OFF cell activations at the spatial locations of luminance edges in the photoreceptor array. ON and OFF cell receptive fields were modeled as 2D difference-of-gaussians (DOG) kernels operating on the photoreceptor array output within each time step. This filtering occurred at four spatial scales. The receptive field sizes were the same for ON and OFF cells. For the smallest scale, the standard deviation of the center gaussian was $\sigma_c = 1.5$ pixel; for the other three scales it was 2.25, 3.3375 and 5.06 pixels. The standard deviation of the surround mechanism was always eight times that of the center mechanism.

ON and OFF cell responses were halfwave rectified to mimic the responses of spiking neurons, which can only be positive. ON and OFF cells respond with different neural gains. The ON cell gain was 0.27 and the OFF cell gain was 1.0. These gain values were based in part on data from achromatic color opponent neurons in the macaque LGN [14, 15], as described in Reference [10].

### Figure 3. Temporal response of the ON and OFF cells. See text for details.

The temporal responses of the ON and OFF cells were identical, except for the difference in overall neural gain. They were modeled as a difference of two scaled beta distributions. The beta distribution is a probability distribution defined on the interval [0,1], whose shape depends on the values of two parameters $p$ and $q$ [16]. “Scaled” versions of these distributions were defined on the interval [0,n] such that the value of the scaled beta distribution at time step $t/n$ is equal to the value of the beta distribution at time $t$. Since a beta distribution is a probability distribution, it is guaranteed to go to zero at $t = 1$. Thus the temporal impulse response of the ON and OFF cells were guaranteed to go to zero by the end of time step $n$. In principle, $n$ could be any number, but it was 30 in the simulations reported here. The first scaled beta distribution was characterized by the parameters, $p = 2$ and $q = 5$ [16]. The relationship between the number of simulation time steps and time in msec was arbitrarily, but a conversion factor of 5 msec/time step is roughly consistent with neural data from cortical responses in cat V1 (see Ref. [17], Fig. 1).

The neural response was halfwave rectified, and the rectified ON and OFF responses went to zero after about 10 time steps of the simulation (see Fig. 3). The maximum value of the rectified response during each 30 time step sensory integration period was “read out” at the next processing stage as the activation level of the ON or OFF cell. The 30 time step integration period thus defined a second important time scale in the model. Except when eye movements were halted to simulated perceptual fading, fixation eye movements occurred at the end of every sensory integration period. Thus, the ON and OFF activations that were read out during each sensory integration period also corresponded to the effective ON and OFF response magnitudes during each fixation period, as seen by the next processing stage. The entire simulation comprised 30 sensory integration periods.

As discussed below, the zero-crossing time of the neuron’s temporal response (see Fig. 3) sets an upper limit on the frequency of the eye movements that can contribute to lightness computation according to the model. This time is estimated from cortical data from anesthesized cat (see Ref. [17], Fig. 1) to be about 50 msec. Thus, the maximum frequency of the relevant eye movements is about 20 Hz, which corresponds to the type of fixational eye movement known as tremor [18].

**ON and OFF Cell Responses Sorted by Eye Movement Direction**

In the model, ON and OFF responses are sorted in visual cortex on the basis of the eye movement direction that generated the response, which is assumed to be encoded as corollary discharge signal. This sorting results in a spatial map of the locations and magnitudes of any ON or OFF activations occurring during each sensory integration period for each eye movement direction.

Since only four eye movement directions were simulated, eight maps were generated, one for each contrast polarity (ON or OFF)/eye movement direction combination. This enumeration assumes that the outputs of the ON and OFF activations at all four spatial scales are combined to produce a single map. If the scales are not combined, there would be 32 neural maps.

The activation maps were computed in object-center coordinates, which is required to generate a cortical representation of perceived surface reflectance, which is the goal of the model. The activation maps were first generated in retinal coordinates, then converted to object-centered coordinates by keeping track of the directions and magnitudes of previous eye movements.

An additional, and important, assumption of the model is that transient ON and OFF activations are triggered only if the temporal response of the ON or OFF cell response was equal to zero at the end of the previous integration period. This ensures that ON and OFF transients will only be generated in response to luminance increments and decrements, respectively, when their receptive fields traverse a luminance edge during an eye movement. As a result of the assumed relationship between the duration of the ON or OFF cell temporal response, and the duration of each fixation (same as a sensory integration period), (see Fig. 3) this condition always held in the simulation. Thus, neural activations were generated after each eye movement at the locations of the luminance edges in the Staircase Gelb display only in the appropriate activation maps.

Fig. 4 illustrates the activations generated in each of the eight maps by the luminance edges in the Staircase Gelb display as the
eye movements traversed the display. Following each eye movement, either the pair of activation patterns corresponding to panels (a) and (e) (leftward eye movement), (b) and (f) (rightward eye movement), (c) and (g) (upward eye movement), or (d) and (h) (downward eye movement) were generated, each with equal probability.

Figure 4. (a)-(d) Locations within the ON activation maps of neural responses to the Staircase Gelb stimulus produced by leftward, rightward, upward, and downward eye movements. (e)-(h) Activations within OFF network maps corresponding to the same eye movement directions as in (a)-(d). Activations have been summed across all four spatial scales of halfwave rectified DOG filter outputs.

Spatial Spreading Within Cortical ON and OFF Networks

At the next stage of the model, the neural responses within the ON activation maps are spatially integrated in a feedforward manner by ON integrator cells possessing large-scale receptive fields; and the neural responses within the OFF activation maps are similarly integrated by OFF integrator cells with receptive fields having the same shape as those of the ON integrator cells. The equation describing the weighting of the responses within the activation maps by the ON and OFF integrator cells is

\[ w(x - x_0, y - y_0, \theta) = \exp \left( -\lambda \log \left( 1 + \frac{(x - x_0)^2 + (y - y_0)^2}{\epsilon} \right) \right) \cos^2 \theta \]  

(1)

where the \([ \cdot ]^+\) denotes halfwave rectification.

Equ. (1) is the product of two separable functions. The first function depends on the distance \(\sqrt{(x - x_0)^2 + (y - y_0)^2}\) between the receptive field center \((x_0, y_0)\) of the ON or OFF integrator cell and the location \((x, y)\) within the activation map of the neural activation being integrated. (see Fig. 6). The second depends on the degree of angular alignment \(\cos \theta\) between the direction of the eye movement that produced the activation and a vector directed from \((x, y)\) to \((x_0, y_0)\).

The receptive field shape of the integrator cells is assumed to be a decaying exponential function with respect to the neuronal layer immediately feeding the cell. When expressed in retinal coordinates as in Eq. (1) and Fig. 6, this exponential profile is modified by cortical magnification. The parameters \(\lambda\) and \(\epsilon\) in Eq. (1) characterize the cortical magnification factor [19]-[22], which is approximated here as an isotropic function of distance from the fovea. The parameter values used in the simulations were \(\lambda = 86, \epsilon = 61.4,\) and \(\rho = 1\).

These feedforward computations result in spatial patterns of activation across two networks of ON and OFF integrator cells. Fig. 7 illustrates the patterns of activity generated within the ON and OFF networks in response to the luminance edges in the Staircase Gelb stimulus corresponding to the four eye movement directions.

Neural Lightness Computation

At the final (output) stage of the model, the neural activations across the networks of ON and OFF integrator cells are summed to produce a neural lightness representation. This process involves two processes in addition to the summation. The first is a network normalization that causes the highest activation level across the lightness network to always equal a constant value that corresponds to the percept of white. This was achieved in the simulation by
shifting all of the pointwise summed activation of the ON and OFF networks by the same amount so that the maximum activation across the lightness network would equal this fixed value.

The second additional process consists of a leaky temporal integration of the normalized activations within the lightness network. This leaky integration causes the lightness percept to gradually develop over time following stimulus onset and to gradually fade when eye movements are halted. If the leaky integration did not exist, the percept of the stimulus would be that of a rapid series of ON and OFF edge induction patterns seen to radiate from the edges in the Staircase Gelb stimulus. There would be no coherent percept of surface reflectance. Hence, the leaky integration serves to bind the ON and OFF inductions generated by different eye movements across across multiple sensory integration periods.

The leaky integration also explains the psychophysical finding that images that are stabilized on the retina fade perceptually over a course of a few seconds [3][4]. To account for the known time course of this perceptual fading, the characteristic time of the leaky integration was assumed to be a few seconds in duration. It should be noted that, in actuality, the characteristic time probably depends in a nonlinear way on stimulus intensity because the rise time of vision is on the order of 100 msec, while fading takes considerably longer. The model could be modified to account for this difference in perceptual rise and fade times, but this was not done here because the assumption of leaky integration as an explanation for perceptual binding is probably an oversimplification, as evidenced by the fact that the perceptual fading of objects that occurs when eye movements are occurs in parts. For example, a corner of square may take longer to fade than other edges of the square [23], suggesting that the underlying cortical computations are based in part on principles of figural organization, and that perhaps the fading time actually depends on the decoherence time of a cooperative neural network that produces this organization. Importantly, however, these complexities do not rule out the basic assumption of the model that the perceptual binding of stimulus features in the process of computing surface reflectance is due in part to a temporal integration of independent ON and OFF network activations generated by luminance edges across separate sensory integration periods.

Predicted lightness matches

The model’s ability to account for quantitative lightness matching data was evaluated by simulating the results of the experiment of Zavagno, Annan, and Caputo, in which observers matched the appearance of standard Munsell chips to that of the papers in their Series A, B, and C. To test the model, the neural activation computed at the center of each paper in the output (lightness) layer of the model was compared to the average perceptual match made to each paper in the psychophysical study.

The fact that eye movement direction is chosen randomly for each eye movement in the model causes the lightness output to also vary randomly. This complicates the process of choosing unique values of the model output to compare to perceptual data. Another complication is the leaky integration that causes the lightness values to vary over time. To avoid these complications, I simplified the model for the purpose of this simulation by simulating the responses of the model to each of the four eye movement directions, then summing this response to generate the model output. Simulated matches were based on the values in the output array evaluated at the square centers. The lightness values computed from this procedure are shown in Fig. 9 along with the actual lightness matches made in the original experiment. As can be seen from the figure, the correspondence between the predictions of the model and the actual matches was quite good. The error was quantified by expressing the least-squares error in the simulated matches as a percentage of the total variances in the psychophysical matches for Series A, B, and C combined. By this measure, the percent error was 2.47%.

Figure 7. Time-course of the simulated lightness of the Staircase Gelb display. Eye movements were halted after Frame 15. The output corresponding to Frame 8 illustrates the Chevreul illusion. See text for further details.

The gradual perceptual onset and fading of the Staircase Gelb image produced by the model is illustrated by Fig. 7, which shows the pattern of activation across the lightness representation after Frames 2, 8, and 30 produced by a simulation in which the eye movements were halted following Frame 15. The simulated appearance of the stimulus after Frame 8, when the percept is peaking, also illustrates the fact that the model reproduces the well-known Chevreul illusion, according to which and ordered series of bands consisting of regions of homogeneous luminance appear scalloped, with the regions within each band that are closer to a neighboring region of higher luminance appearing relatively darker than other regions with the same luminance band [2].

Figure 8. Simulated lightness matches in response to Series A, B, and C of Zavagno et al. plotted along with the average lightness matches made by human observers (red, green, and yellow symbols). Blue symbols indicate the simulated matches produced by the model when the Staircase Gelb stimulus was presented against a white background.
Extension to Color Vision

In principle, the neural lightness model described above could be extended to color vision by including a complete set of “double-opponent” neurons (Fig. 8) as edge detectors in the neural edge integration model. The resulting edge integration model would differ from the classic Retinex color vision model of Land and McCann [25] in that the additional chromatic dimensions of color space would be introduced as a property of the neurons performing edge detection, rather than at the level of the photoreceptors. However, to do this in a manner that accurately accounts for quantitative psychophysical color matching data remains an unsolved problem and is the subject of my current research.

Figure 8. Double-opponent receptive fields. In addition to chromatic ON and OFF cells, two other four other classes of cortical neurons combine the output of the S, M, and L photoreceptors in two additional ways.

References


Author Biography

Michael Rudd earned B.S. degrees in physics and psychology from the University of California, Davis (1978) and a Ph.D. in experimental psychology from the University of California, Irvine (1987). He did post-doctoral research in neural networks with Stephen Grossberg at Boston University. While on the faculty at Johns Hopkins University, he conducted pioneering work on stochastic neural models of decision making and adaptation. He has taught at the University of Washington and worked as a Research Scientist for the Howard Hughes Medical Institute. He is currently a member of the Psychology and Neuroscience faculties of the University of Nevada, Reno and a manager of the Computational Modeling and Analysis Core of the NIH-funded Center for Integrative Neuroscience at UNR. His research combines computational neuroscience with psychophysical investigations of lightness, color, and human rod and cone vision.