Linking perceived to physical contrast: Comparing results from discrimination and difference-scaling experiments

Christopher Shooner

McGill Vision Research, Department of Ophthalmology & Visual Sciences, McGill University Montreal, Quebec, Canada ✉

Kathy T. Mullen

McGill Vision Research, Department of Ophthalmology & Visual Sciences, McGill University Montreal, Quebec, Canada ✉

Psychophysical approaches that allow us to estimate how perceived stimulus intensity is linked to physical intensity are import tools for studying nonlinear transformations of visual signals within different visual pathways. Here, we investigated how stimulus contrast is encoded in achromatic and chromatic pathways using simple grating stimuli. We compared two experimental approaches to this question: contrast discrimination (increment detection thresholds measured on contrast pedestals) and the maximum likelihood difference scaling (MLDS) approach introduced by Maloney and Yang (2003). The results of both experiments are expressed using simple models that include a transducer function mapping physical contrast to an internal signal the observer uses in making judgments, and an estimate of the variability of this representation (internal "noise"). We found that the transducers derived from both experiments have a similar form, but occupy different ranges of physical contrast in different stimulus conditions, reflecting difference in contrast sensitivity. This is consistent with past discrimination results, and in the difference-scaling case provides new evidence supporting the idea that suprathreshold chromatic and achromatic contrast are processed similarly, once differences in contrast sensitivity are taken into account. Model estimates of internal noise were higher in the difference-scaling experiment than the discrimination experiment, a finding we attribute to a difference in task complexity. Finally, we fit an alternative version of the MLDS model in which internal noise increased with response level. This alternative was no better at predicting holdout data in a cross-validation analysis than the original constant-variance model.

Introduction

How the perceived intensity of a physical stimulus relates to its physical intensity is a foundational question in psychophysics that remains incompletely answered. This study aimed to compare two experimental approaches that have been applied to this problem: contrast discrimination and contrast difference scaling. The fact that physical and perceived intensity are not related by a simple linear correspondence follows from Weber's law: as intensity increases, larger differences in intensity are required for two stimuli to be perceptually discriminable. Fechner pointed out that this pattern would be expected if a compressive nonlinear transformation existed between the physical and perceptual: when higher-intensity inputs are compressed to a smaller range of outputs, the same physical increment has smaller perceptual consequence (Fechner et al., 1860/1966).

Many subsequent works in the domain of contrast discrimination have expanded on this idea. (Foley, 1994; Georgeson & Meese, 2004; Georgeson & Meese, 2006; Goris, Wagemans, & Wichmann, 2008; Goris, Putzeys, Wagemans, & Wichmann, 2013; Klein, 2006; Kontsevich, Chen, & Tyler, 2002; Legge & Foley, 1980; Meese, 2004; Olzak & Thomas, 2003; Pelli, 1985) Several of these studies developed full models of contrast processing to explain observed threshold-versus-contrast functions. The central feature of such models is a nonlinear transducer: a mapping from physical stimulus intensity to an internal signal that the observer uses in making discrimination judgments. To link these models quantitatively to behavioral data (e.g., proportion correct in a forced-choice discrimination task) an additional model component is required which describes the variability of the internal contrast representation. This variability or "noise" is often modeled as a normally-distributed stochastic signal added to the transducer output. In this case standard signal-detection theory methods can be used to predict the probability of an observer giving one response over another.

The "compressive transducer plus noise" model can be applied to any discrimination data showing a steady increase in threshold with contrast. It leaves open the question, however, of how well a transducer modeled on threshold-level performance matches our subjective percept of stimulus intensity. This question is complicated by the fact that the form of the modeled transducer depends on the model's assumptions regarding internal variability. As pointed out by multiple studies, because performance depends on a signal-to-noise ratio, it is impossible in most cases to use a simple discrimination experiment to jointly constrain the form of a transducer (the signal) and the form of the noise (but see Solomon, 2007; see Georgeson & Meese, 2006 for review and discussion) (Georgeson & Meese, 2006; Kingdom, 2016; Klein, 2006; Kontsevich et al., 2002). If instead of assuming constant-variance noise, noise is modeled as level-dependent, increasing with stimulus contrast as single neurons in visual cortex are known to do (Goris, Movshon, J. A., & Simoncelli, 2014), an equally good model fit could be obtained with a transducer having a very different shape (Kingdom, 2016; Kontsevich et al., 2002).

Perceptual scaling experiments treat the question of linking perceived to physical contrast with a different approach (Aguilar, Wichmann, & Maertens, 2017; Brown, Lindsey, & Guckes, 2011; Maloney & Yang, 2003; Whittle, 1992). In a contrast-difference scaling experiment, observers are presented with pairs of stimuli differing in contrast and judge how different the two appear from each other (Knoblauch, Marsh-Armstrong, & Werner, 2020; Kulikowski, 1976; Whittle, 1992). Unlike contrast discrimination, this is a subjective judgement comparing stimuli with very different intensity that are therefore highly discriminable. This judgment can be framed in a two-alternative forced-choice experiment by presenting two contrast intervals and asking the observer which pair creates a larger difference in apparent contrast. The maximum-likelihood difference scaling (MLDS) framework introduced by Maloney and Yang (2003) pairs this experimental approach with a model of how observers judge these differences. Like the models described above, this includes a transducer that maps physical contrast to perceived contrast, and an estimate of the variability in the transducer output. Again using a signal-detection model, this offers a prediction of the probability that the observer will judge one interval larger than the other. Comparing this prediction to observed responses allows a maximum-likelihood fitting procedure to find the best model transducer and modeled internal noise to match the observed responses (Knoblauch & Maloney, 2008; Maloney & Knoblauch, 2020; Maloney & Yang, 2003). The MLDS approach is not limited to the contrast dimension and has also been used successfully to derive perceptual scales of hue (Maloney & Yang, 2003), test theories of categorical color judgement (Brown et al., 2011), evaluate perceived surface gloss (Obein, Knoblauch, & Viéot, 2004), as well as other uses (Aguilar et al., 2017; Knoblauch et al., 2020).

The discrimination and difference-scaling approaches use very different experimental methods but offer models of the same form: a nonlinear mapping from physical contrast to an internal contrast representation, and an estimate of the variability of that representation. Here, we ask how well the results of the two methods agree if they are tested in parallel under the same conditions. With a similar goal, Devinck and Knoblauch (2012) reported good agreement between discrimination and MLDS results in their study of the watercolor effect. Aguilar, Wichmann, and Maertens (2017) also investigated this relationship for selected stimulus conditions, measuring perceived tilt from texture, and found good agreement between directly-measured tilt-discrimination thresholds and those derived from an MLDS approach. A more detailed description of the MLDS method can be found in Maloney and Knoblauch (2020), together with a review of its applications and a discussion of its relationship to previous scaling methods.

In this study we investigated the nonlinear mapping from physical to perceptual using a more low-level stimulus attribute: the contrast of simple grating stimuli. We measured contrast discrimination thresholds over a range of "pedestal" contrasts. With the same stimuli we performed a difference-scaling task and fit the results with the MLDS model. We used five different grating stimuli to determine whether this mapping differs across different conditions and to determine whether the agreement between the two methods depends on the stimuli selected. Stimuli were designed to bias the responses toward different neural pathways and included both chromatic and achromatic contrast. The two chromatic stimuli activated both the L/M and S cone opponent pathways, respectively (Mullen & Losada, 1994). Three achromatic stimuli were chosen with spatiotemporal parameters designed to span the range of preferential activation between magnocellular and parvocellular visual pathways. For all stimulus types, the modeled transducers derived from both experiments showed the expected compressive shape, but the modeled internal noise level was significantly higher in the MLDS experiment, across all observers, revealing an important difference between the approaches.

Methods

Participants

Four participants (three female) served as observers: one author and three individuals unaware of the hypothesis being tested. All procedures were approved

Experiment 1: Contrast discrimination

A
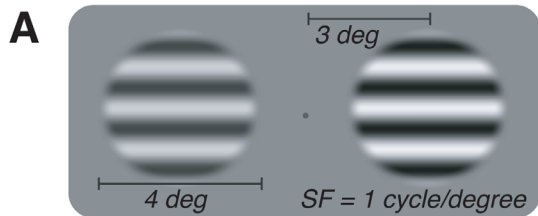
"Which stimulus has the higher contrast?"

Experiment 2: Contrast difference scaling

B

*"Which of the bottom stimuli differs most
from the stimulus on top?"*

Figure 1. (A) Two gratings were shown to the left and right of fixation for one second with constant spatial phase. The observer reported the left or right grating as higher in contrast. (deg: degree, SF: spatial frequency.) (B) Three gratings, identical to those used in the first experiment, were presented at the same eccentricity. The observer reported which test (left or right) differed more from the reference (top) in apparent contrast.

by McGill University's Institutional Review Board and conformed to the Declaration of Helsinki.

Visual stimuli

Horizontal sinusoidal gratings were presented in 4° circular apertures, the outer 1° smoothed with a raised-cosine profile. In the discrimination experiment, two stimuli were presented 3° to the left and right of a central fixation point. In the scaling experiment, three stimuli were presented at 3° eccentricity with polar-angle spacing of 120°, the reference stimulus above fixation and the two test stimuli in the lower left and right fields (see Figure 1). Spatial phase was constant, identical for all stimuli within a trial, and

randomized across trials. Stimuli were presented for one second with contrast ramped up and down over the first and last 200 ms by a raised cosine. In one achromatic condition and both chromatic conditions, gratings had a spatial frequency of 1 cycle/degree (c/deg) with no temporal modulation. Two additional achromatic conditions were tested with more extreme temporal or spatial frequencies in order to activate magnocellular and parvocellular pathways preferentially. The "M-biased" stimulus was lower in spatial frequency (SF = 0.5 c/deg) and contrast modulated at 8 Hz. The "P-biased" stimulus was higher in SF (8 c/deg) with no contrast modulation.

Stimulus chromaticity was defined within a cone contrast space as a modulation through the origin in a direction defined by a triplet { *l, m, s* }, representing the fractional change in excitation of the long-, medium-, and short-wavelength-sensitive cones. Stimulus contrast was defined as the depth of this modulation, expressed as a vector length $c = \sqrt{l^2 + m^2 + s^2}$. This definition differs from Michleson contrast; a full-contrast achromatic grating modulates each cone by 100% (*lms* = [1,1,1]) and so has a total cone contrast of $\sqrt{3} \cong 173\%$. Isoluminant chromatic stimuli were designed to isolate post-receptoral L/M and S/L+M cone opponent responses, referred to as *red-green* and *blue-yellow*, respectively. For the blue-yellow, this is simply an S-cone-isolating stimulus, as S cones are not thought to contribute significantly to luminance or red-green mechanisms (Mullen & Losada, 1994). The red-green (RG) stimuli were defined as *lms* = [1, −*a*, (1−*a*)/2], with the value of *a* chosen separately for each observer to make the stimulus isoluminant, based on a preliminary minimum-motion experiment (Anstis & Cavanagh, 1983). The S-cone component ensured that S − (L+M)/2 = 0, nominally eliminating any blue-yellow signal from this stimulus.

Contrast discrimination

We measured thresholds for detecting contrast increments added to a range of baseline "pedestal" contrasts. Two gratings were displayed to the left and right of fixation, one at the pedestal contrast level, the other with a variable contrast increment added to the pedestal (Figure 1). The observer's task was to report the higher contrast in a forced-choice procedure via a button press. Increment size was varied using an adaptive staircase. A block of trials tested one stimulus type and one pedestal contrast. Within a block, two independent staircases (2-down/1-up) were randomly interleaved. The block terminated after five reversals of both staircases. Each condition was repeated in six to ten blocks and the resulting 200 to 500 trials were pooled and fit with a Weibull psychometric function to extract a threshold, defined as the contrast yielding 75% correct responses. In a bootstrap procedure we fit

each threshold 100 times to different randomly-sampled subsets containing 90% of the trials. We report the median of these bootstrap distributions. Pedestal contrasts were defined as multiples of detection threshold, separately for each observer. We tested 10 to 12 pedestal levels with octave spacing, ranging from $0.125\times$ threshold to the highest contrast possible, which varied across conditions (typically $32\times$ or $64\times$ threshold). In some cases, half-octave spacing was used to give greater precision.

Descriptive model of discrimination data

We fit each dipper function with a descriptive model, in which threshold is constant at low contrast ($t = t_0$), follows a power law at high contrast ($t = \alpha c^p$), and smoothly transitions between these regimes following a sigmoidal weighting function:

$$t = (1 - w)\, t_0 + (w)\, \alpha c^p$$

$$w = \frac{c^\beta}{c_0^\beta + c^\beta}$$

This transition is not meant to describe a visual mechanism but allows us to isolate the suprathreshold range of interest by defining a "transition contrast" (c_t) where the shift to power-law behavior is 95% complete ($w = 0.95$).

Contrast-difference scaling

Our second experiment used the MLDS framework introduced by Maloney & Yang (2003), in which a model is used to derive a perceptual scale representing a physical stimulus attribute. The experiment itself did not require observers to scale stimuli. A two-alternative forced-choice procedure was used to measure perceived differences in contrast between pairs of gratings. Observers were presented with three stimuli: a reference and two tests, one lower in contrast than the reference and the other higher (Figure 1). Observers were asked to compare both tests to the reference in terms of apparent contrast and report which contrast *difference* was larger, test 1 versus the reference or test 2 versus the reference. Reference and test contrasts were chosen from a fixed set of six to ten values, which had a spacing of approximately four threshold units as determined from our discrimination experiment (detailed in Results). Referring to contrasts by their place in this series, we constructed a list of all possible ordered triplets for which the contrast differences of the two tests from the reference differed by no more than one contrast step. For example, the third, fifth, and eighth contrasts comprised a valid triplet, {3,5,8}, but {3,5,9} was excluded, because the size of the first interval was two contrast steps, and the second interval was four.

This led to 14 triplets in the minimum case of six tested contrasts, and 52 triplets when 10 contrasts were tested. As quantified below, this set included triplets that presented the subject with both easy and difficult judgments. Within a block, all triplets were presented in random order, with the position of the lower-contrast test (left or right) also randomized. The observer's task was to report the left or right interval as the larger contrast difference. Each triplet was presented 20 times or more across multiple blocks. Thus between 280 and 1040 trials contributed each model fit described below, depending on the number of contrasts that could be tested.

MLDS fit to scaling data

We fit our contrast scaling data with the signal detection model introduced by Maloney and Yang (2003). Details of the model and fitting procedure have been presented elsewhere (Knoblauch & Maloney, 2008; Maloney & Knoblauch, 2020; Maloney & Yang, 2003). The model describes a transducer, $\psi(c)$, which maps a physical stimulus property (in this case contrast, c) to an internal representation that the observer uses in making judgments, commonly referred to as a perceptual scale. In the case of our triplet experiment, we assume that each of the three grating contrasts *c1, c2, c3,* elicit internal responses $\psi(c1)$, $\psi(c2)$, $\psi(c3)$, and the observer selects the second contrast interval as larger if $\psi(c3) - \psi(c2)$ is greater than $\psi(c2) - \psi(c1)$. These internal responses are assumed to be stochastic, and the resulting uncertainty in the decision is modeled by assuming constant-variance Gaussian noise added to each value. The probability that the second interval is chosen is then derived from the signal detection model described by Maloney and Yang (2003). The model is parameterized by assigning a discrete ψ value to each of the contrasts tested. We normalize the scale by assigning ψ values of 0 and 1 to the lowest and highest tested contrasts, and fit an estimate of internal noise (σ) as an additional parameter, leading to *n-1* free parameters for *n* tested contrasts. Parameters were adjusted in an optimization routine (*fmincon* in MATLAB) to maximize the total likelihood of the data set over all trials. In a bootstrap procedure we repeated the fit 100 times in each condition using a randomly sampled 90% of trials. We report the median ψ and σ values obtained from bootstrap distributions.

Apparatus

Stimuli were generated in MATLAB (The MathWorks, Natick, MA, USA) using the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli & Vision, 1997) and displayed on a CRT monitor (DiamondPro 2070;

Mitsubishi Electric Corporation, Tokyo, Japan) with a mean luminance of 50 cd/m^2 and chromaticity xy = {0.31, 0.33}. Observers viewed the screen from a distance of 80 cm, at which it subtended 22° × 27° of visual angle. A Bits# visual stimulus generator (Cambridge Research Systems, Kent, UK) was used to control the amplitude of each color channel with 14-bit precision. Nonlinearity in the output of each color channel was characterized using a SpectroCAL spectroradiometer (Cambridge Research Systems) and corrected in software. The measured emission spectra of the monitor phosphors were integrated with psychophysically derived cone fundamentals (Smith & Pokorny, 1975) to create a linear transformation specifying the RGB values required to elicit any target triplet of cone excitation levels.

Results

Contrast discrimination

Contrast-increment thresholds measured over a range of pedestal contrast levels showed a typical "dipper" shape (Legge & Foley, 1980), as shown in Figure 2 for one subject and one condition (achromatic). Thresholds were reduced on low-contrast pedestals and at higher contrast increased steadily with pedestal level. The high-contrast arm of the dipper, linear on this double-logarithmic plot, suggests a compressive power-law relationship: thresholds increased proportionally to pedestal contrast raised to an exponent less than 1 (0.64 in this case). We used a descriptive model (see Methods) to define this power-law range of contrasts and estimate the exponent (slope) separately for each dipper function.

In this study we are interested only in suprathreshold contrast processing so we restrict further analysis to this power-law range of contrasts. In this range, discrimination can be described with a simple transducer model: a mapping from physical contrast to an internal response the observer uses in making judgments (Foley, 1994). We assume a pair of stimuli are discriminable when they elicit responses differing by some criterion amount. A compressive transducer such as that in Figure 2A has shallower slope at high contrasts, and a larger difference in physical contrast is required to achieve the same difference in output (threshold is higher). This can be formalized using signal detection theory by assuming that responses are stochastic. When constant-variance Gaussian noise is added to the transducer output, the criterion difference in responses is defined as one standard deviation of the noise ($d' = 1$).

The transducer output is expressed in units of noise standard deviations (σ). The slope of this curve
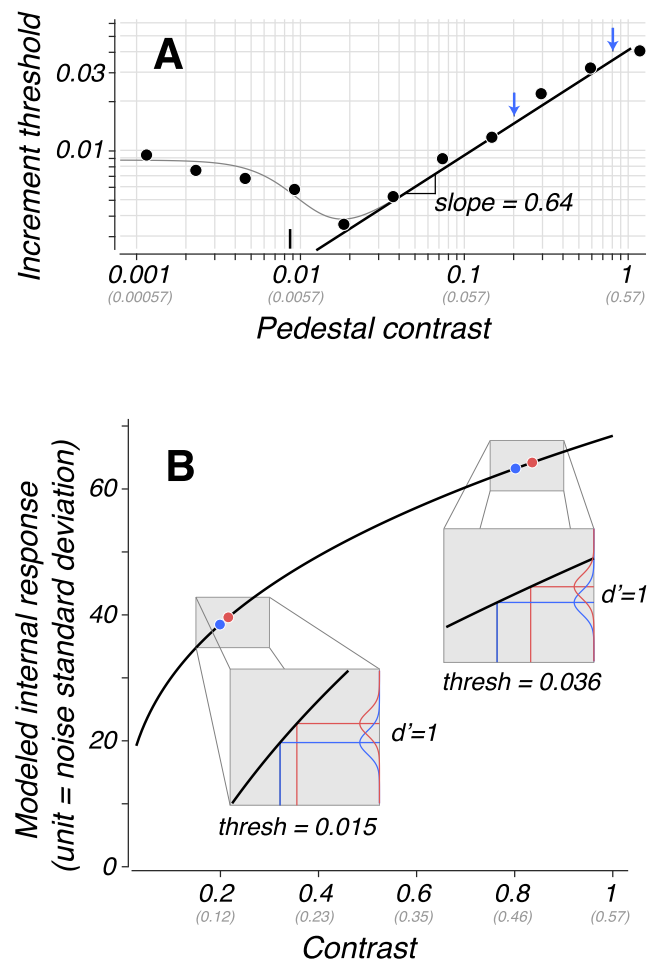

Figure 2. (A) For one subject and condition (achromatic), increment detection thresholds are plotted with respect to pedestal contrast, both in units of total cone contrast (see Methods). Michelson contrast values are given in parentheses. The thin curve shows a descriptive model fit which includes a straight-line section at high contrast, capturing a power-law relationship (thick line). The vertical tick represents threshold for detecting the stimulus with no pedestal. (B) A model of a saturating transducer is derived from the power-law portion of the dipper function (and valid only in that range). For two example pedestal levels (*blue arrows* in A and *blue dots* in B) the gray inset panels show that a larger increment is required at the higher pedestal level to achieve outputs separated by 1 standard deviation of constant-variance noise ($d' = 1$).

is then in units of d' per unit contrast, equivalent to contrast sensitivity (1/threshold). We obtain the transducer by integrating the inverse of our power-law model describing threshold, which leads to another power-law description for the transducer: $r = a*c^b$. For comparison with our next experiment, it is important to note that observers are quite good at fine contrast discrimination. For example, at a pedestal level of 80% cone contrast (46% Michelson contrast; see Methods), the increment threshold was 3.6% in this example case,
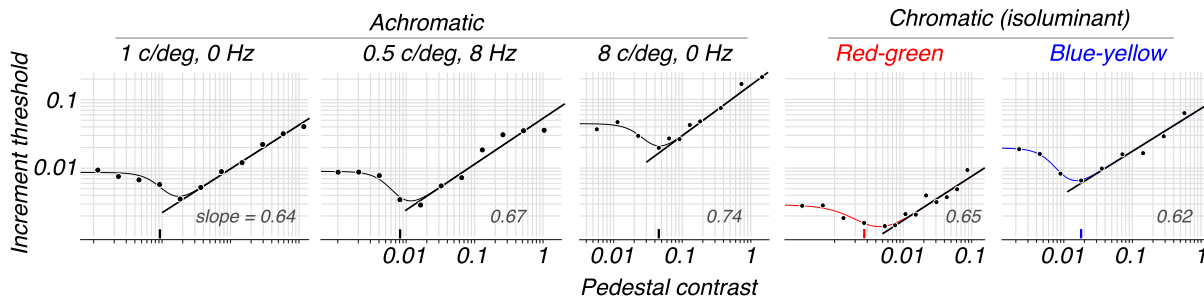
Figure 3. Threshold-versus-contrast (dipper) functions are shown for one subject in five different conditions differing in chromaticity and spatial and temporal frequency. Despite large differences in sensitivity across conditions, dipper functions showed a similar form. *Vertical tick marks* on the horizontal axis represent detection thresholds. *Thin curves* show fits of a descriptive model described in Methods. *Thick black lines* highlight the portion of the dipper exhibiting a power-law relationship. The slopes of these lines (exponents of the power law) are shown in the lower right of each panel.

less than a 0.5% relative increment. This is reflected in the model as a high signal-to-noise ratio, seen in Figure 2B where the transducer spans a range of 70 σ units.

We repeated the discrimination experiment using achromatic stimuli with high temporal frequency (TF) or spatial frequency (SF) content, as well as isoluminant chromatic stimuli (red-green and blue-yellow; see Methods for definition). The high-TF stimulus was presented with low spatial frequency (0.5 c/deg) and a sinusoidal contrast reversal (flicker) at 8 Hz. The high-SF stimulus was 8 c/deg with no flicker. Chromatic stimuli were presented with the same spatiotemporal structure as our baseline achromatic condition described first. The high-TF and high-SF conditions were designed to bias internal responses more toward magnocellular and parvocellular pathways, respectively. Figure 3 shows contrast discrimination

results for these five conditions for one observer. These dipper functions occupy different ranges of absolute contrast, as expected from known differences in contrast sensitivity across conditions: Thresholds for high-spatial-frequency stimuli were elevated; thresholds for red-green stimuli were significantly lower than all others (Bird, Henning, & Wichmann, 2002; Mullen, 1985; Mullen & Losada, 1994). The form of the dipper functions was similar, however, with a "dip" near detection threshold (marked by a vertical tick in each plot), and a power-law increase in threshold at higher contrast, with similar slope across stimuli.

Figure 4 shows dipper functions for all observers and all conditions, on absolute cone-contrast axes (top row) and threshold-normalized axes (bottom row). Normalizing by detection threshold brought most dipper functions into good alignment, consistent
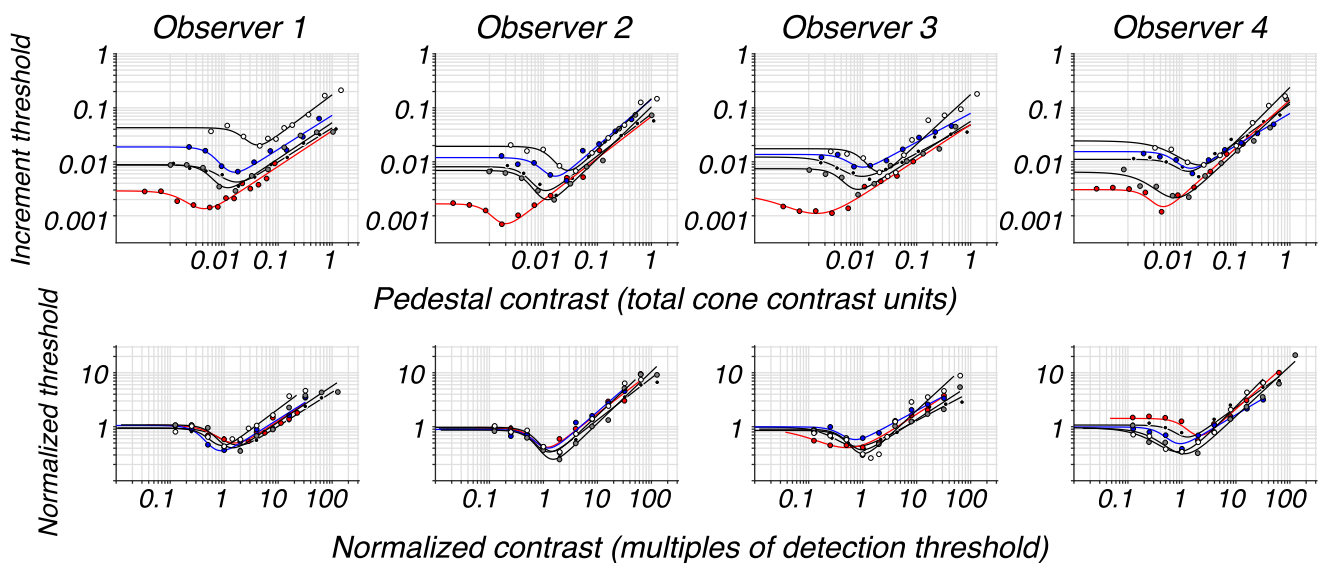

Figure 4. Threshold versus pedestal contrast is plotted for four observers (columns) with all conditions overlaid. The *top row* expresses this in units of absolute cone contrast. In the *bottom row* both axes are in normalized units derived by dividing contrast by detection threshold, separately for each condition.
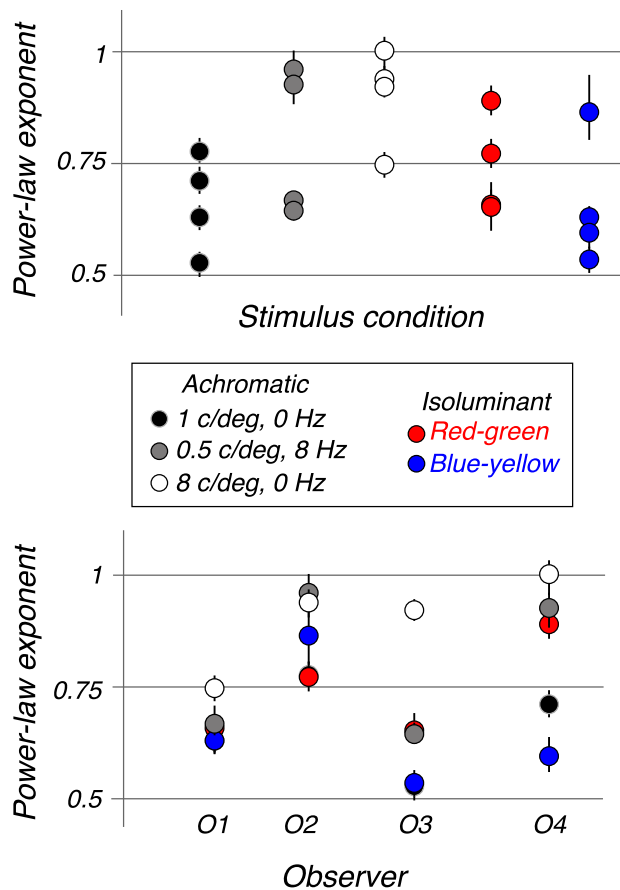
Figure 5. Power-law exponents derived from discrimination results (slope of dipper functions) are shown for all stimuli and all observers. (A) Exponents grouped by stimulus with one point per observer. (B) The same data as in A but grouped by observer. *Error bars* represent 95% confidence intervals from a bootstrap procedure.

with previous findings (Switkes, Bradley, & DeValois, 1988).

Power-law exponents ranged from 0.5 to 1, but did not show a clear dependence on stimulus type or observer, with the possible exception of steeper slopes for Observer 2 (see Figure 5).

Maximum-likelihood difference scaling (MLDS)

We used the results of our discrimination experiment to choose a series of suprathreshold contrasts for use in the MLDS experiment, separately for each subject and condition. These were approximately equally spaced perceptually in the range where discrimination showed power-law behavior. Specifically, starting at the transition contrast (c_t) defined above, we used the power-law model to estimate the increment threshold at this pedestal level. We added four times this value to c_t to obtain the second contrast in the series. Using a new threshold value estimated at this second contrast we

obtained a third value in the same way and repeated this until we obtained 10 contrasts or reached the maximum possible contrast. Thus, any two adjacent contrasts were approximately four threshold units apart. The spacing of contrasts is an important design question for MLDS, as it influences how confidently the observer can make the required judgments. At one extreme, very difficult judgments would lead to chance performance across all triplets and provide no information. At the other extreme, if all judgments were trivially easy, the model would not be able to estimate internal variability. To confirm that we fell between these extremes we computed for each triplet the fraction of trials in which an observer reported the second interval as larger. Grouping across all observers and conditions, we found the distribution of this empirical probability (not shown) to be quite flat, with 58% of conditions falling in the bottom and top quartiles ($p < .25$ or $> .75$). Thus there was adequate variability overall, with many conditions led to highly consistent responses.

The results of the MLDS fit are shown in Figure 6 for Observer 1 in the achromatic condition (as in Figure 2). The free parameters of the model include a scale value for each tested contrast, with the first and last fixed at 0 and 1. An additional parameter, σ describes the standard deviation of constant-variance Gaussian noise added to each output. For three example contrasts, Figure 6 shows the modeled distributions of internal responses, and diagrams how these are compared to derive a decision variable estimating which contrast difference is larger.

The model transducer obtained from MDLS in this case had a compressive shape similar to that obtained from the discrimination experiment. We found, however, that the modeled internal noise was larger for MLDS than for discrimination. To compare the two experiments, we scaled the discrimination model to span the same range as the MLDS model. The first transducer predicts discriminability based on the *difference* between two responses, relative to the noise level σ. It is therefore not affected by an additive offset, and can also be scaled arbitrarily given that response amplitude and noise level are scaled identically. We shifted and scaled the transducer curve to have values of 0 and 1 at the lowest and highest contrast tested in the MLDS experiment. The results are shown in Figure 7 for the same example condition as Figure 6. After scaling, the discrimination model had a σ less than half that of the MLDS model. Note that the σ values we report represent the variability of the transducer output and not that of the decision variable. Given the noise model used here, differencing two response variables leads to a distribution with twice the variance of the inputs. As MLDS relies on a difference of differences, the decision variable has four times this variance (see Figure 6), or a standard deviation of 2σ. For discrimination (comparing just two stimuli) this
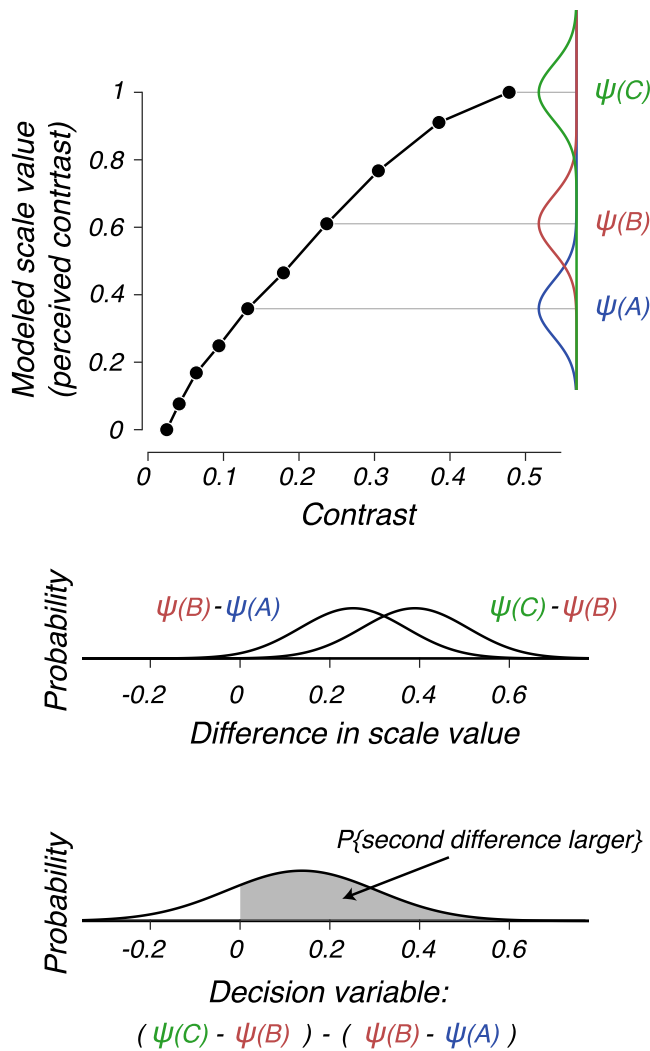
Maximum-likelihood difference scaling (MLDS)

Figure 6. The MLDS model: each contrast level is mapped to an internal scale value (psi) which is stochastic and normally distributed with standard deviation sigma. Distributions of psi elicited by three example contrasts are shown in red green and blue. Psi values are compared to estimate the sizes of the contrast intervals A-B and B-C. These intervals are compared to give a noisy decision variable, with positive values indicating B-C is larger than A-B. The portion of the distribution above zero represents the probability that the second interval is chosen. Psi and sigma values are adjusted to maximize the likelihood of observed reports given this model prediction.

would be $\sqrt{2}\sigma$. This difference is potentially related to a difference between the tasks, as discussed later, but mathematically it does not affect our fitted σ values, given the model assumptions.

Transducers derived from both experiments are shown in Figure 8 for all observers and conditions, in the same format as Figure 7. Independent of modeled noise level, we also compared the shape of transducers using a "compression index," defined as the level of the
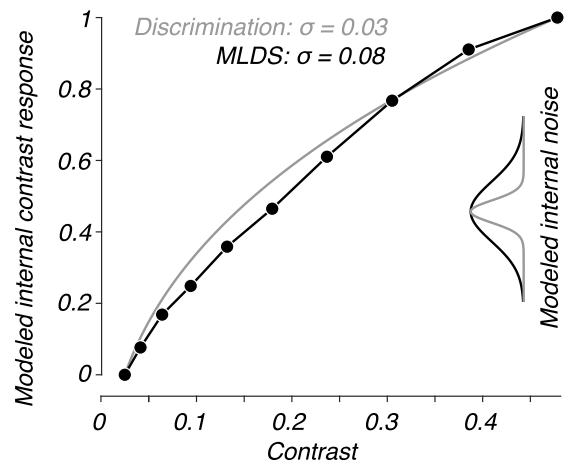
Figure 7. Transducer models derived from contrast discrimination (*gray*) and MLDS (*black*). The discrimination model was shifted and scaled to match the range of MLDS (0-1). The internal noise level (*sigma*) for discrimination, initially fixed at 1 (see Figure 2) was scaled down by the same factor as the transducer. The resulting *sigma* was less than half that derived from the MLDS model.

normalized curve at the halfway point between lowest and highest contrast. For a linear curve this would be 0.5; values greater than 0.5 indicate compressive nonlinearity. Where the MLDS experiment used an even number of contrast levels (and so lacked one at the halfway point) we used linear interpolation for this analysis. Figure 9 compares the sigma values and compression indexes between the two experiments, for each observer and condition.

Modeled internal noise was higher for MLDS than discrimination in all cases. We did not find a systematic difference across stimulus conditions in terms of noise level, transducer shape, or the degree of agreement between the two experiments. An exception is found in the high spatial frequency case, where for two observers the MLDS results show a more linear transducer compared to that from discrimination and compared to other conditions. Observer 2 also showed this effect but to a much smaller extent, and unfortunately Observer 1 did not complete this condition. So while this suggests a possibly interesting effect, we do not have the power to draw a strong conclusion. In some cases, transducers were more compressive in the achromatic than chromatic cases, especially in the low-spatial-frequency flicker condition, but this was not found consistently across observers. We did find several clear differences between observers, independent of stimulus type. While noise estimates from discrimination were highly consistent across observers and conditions, those from MLDS were higher for Observer 4 than other observers across all stimulus types. The shape of the transducers from MLDS were consistently more linear for Observer 3
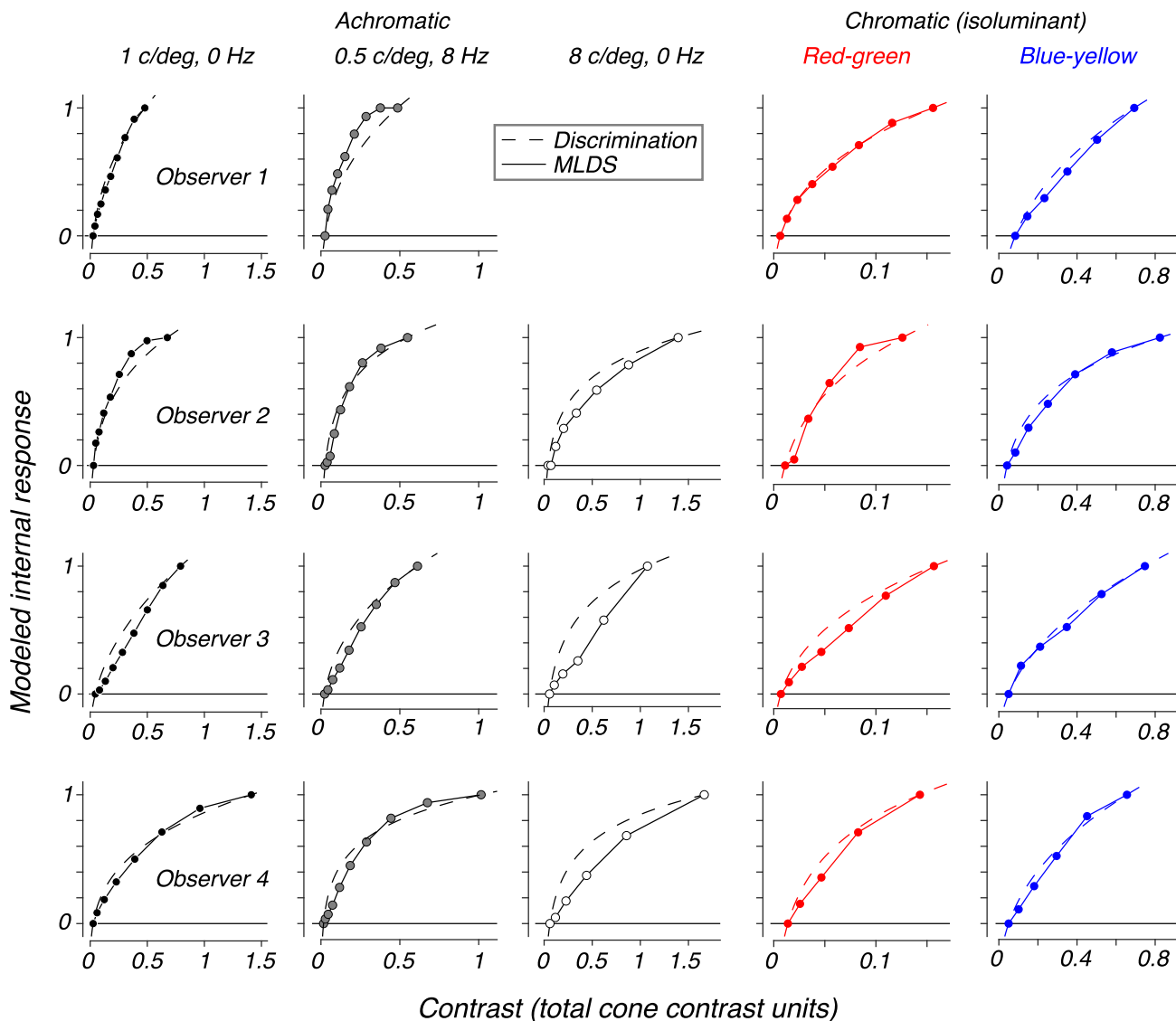
Figure 8. Model transducers for all observers and conditions, in the same format as Figure 5. Both transducers had a compressive shape across all stimuli. Modeled internal noise was larger for MLDS than discrimination in all cases. See Figure 9 for comparison across conditions. Note that Observer 1 did not perform the high-spatial-frequency condition due to high detection threshold.

compared to other observers, and in comparison to that observer's transducers from discrimination. Overall, we found the results of the contrast discrimination experiment to be more consistent across observers and stimulus conditions, while MLDS results were more variable (see rightmost panels of Figure 9).

The range of physical contrasts tested in the scaling experiment varied widely across stimulus condition, as each contrast series was tailored to the stimulus based on discrimination results. To further compare the shape of model transducers, we expressed them on a contrast axis normalized to sensitivity, defined as multiples of detection threshold. Figure 10 replots the MLDS results both on absolute and normalized axes. For the first three observers, this normalization brings the results from all stimuli into alignment,

and highlights the finding that Observer 3 showed a more linear curve across all stimuli. Observer 4's results do not align on normalized axes. The cause of these individual differences is unclear, but it is possible that observers adopted different strategies in making subjective contrast comparisons, and that Observer 4's strategy even varied across conditions. This is consistent with the fact that model estimates of internal noise were higher and more variable for Observer 4 than others.

Alternative MLDS model with level-dependent noise

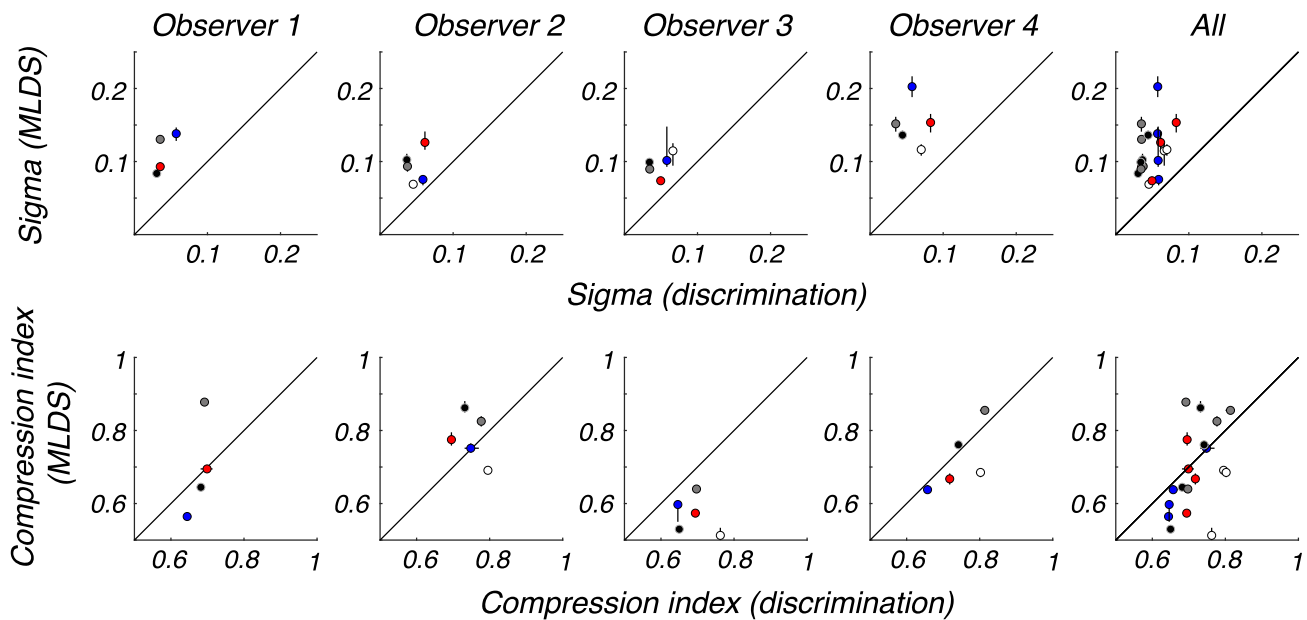The models used thus far to describe both our experiments assume constant-variance Gaussian

Figure 9. Each transducer was described by an internal noise parameter *sigma* and a compression index, defined as the value of the curve midway between lowest and highest contrast (0.5: linear, >0.5: compressive). Noise estimates (*top row*) were higher for MLDS than discrimination in all cases, with no clear dependence on stimulus type. Values were consistent across subjects and conditions with the exception of Observer 3, who showed larger sigma values only in the MLDS experiment. Values of the compression index (*bottom row*) also showed no clear stimulus dependence but systematic individual differences, with Observer 2 showing more linear MLDS results compared to discrimination and compared to other observers. The rightmost panels show all observers together to highlight that discrimination is more consistent than MLDS by both measures. *Error bars* represent 95% confidence intervals from a bootstrap procedure.

variability; the noise level does not vary with contrast or response. There is debate as to whether this noise model is appropriate for describing perceptual variability (Georgeson & Meese, 2006; Kingdom, 2016; Klein, 2006; Kontsevich et al., 2002; Solomon, 2007), fueled in part by the noise properties of visual neurons, which show level-dependent variability with response variance increasing with mean response level (Goris et al., 2014). To address this in the MLDS case, we fit an alternative model, in which variance increased linearly with scale value (ψ). This noise model, $v = v_0 + \alpha\psi$, has two free parameters: the variance at the lowest contrast tested (v_0) and a proportionality constant (α) controlling the increase in variance with mean level. These replace the single σ parameter of the original model, and otherwise the model structure and fitting procedure are unchanged. We compared this model to the original using a cross-validation procedure. Randomly selecting 10% of trials as "holdout" data, we fit both models to the remaining 90% of the data, and evaluated the models' ability to predict the holdout data by computing the likelihood of the data given the model parameters. We found that both versions of the MLDS model predicted holdout data equally well, as shown in Figure 11. Moreover, the fitted values of the parameter α were very small (<0.1), compared to typical Fano factor values of 1 or greater. This was a

consistent result, found even when starting parameters given to the fitting algorithm were greater than 2, and suggests that the optimization routine tended toward the constant-variance case.

Discussion

This study addressed two important questions about how perceived contrast is linked to physical contrast. First, we asked whether the internal representation of contrast differs across the distinct visual pathways that are found in the human visual system. Second, when these representations of perceived contrast are estimated using two different approaches, contrast discrimination and MLDS, how well do the outcomes agree?

We performed both the discrimination and difference-scaling experiments using a range of grating stimuli designed to activate different visual pathways. Color vision is thought to rely on three separate mechanisms that differently combine signals from the three cone types (Cole, Hine, & McIlhagga, 1993; Sankeralli & Mullen, 1996; Sankeralli & Mullen, 1997). The first of these represents luminance by summing cone inputs; two cone-opponent mechanisms compare
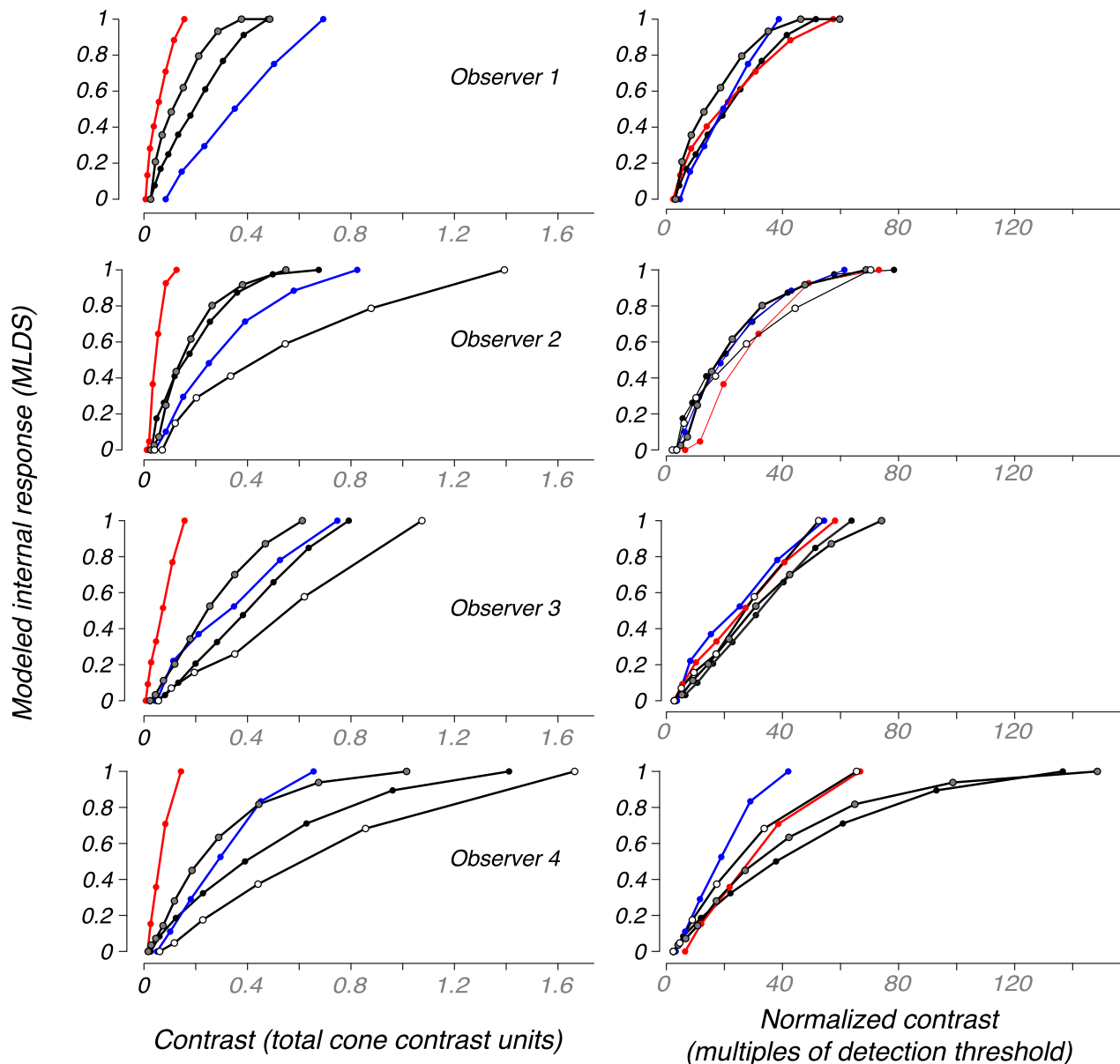
Figure 10. MLDS results are plotted for each observer and condition, both on absolute contrast axes (left column) and a normalized axis defined as multiples of detection threshold. For three observers, normalizing by threshold brought MLDS results into alignment. See Figure 5 for the color code defining stimulus conditions.

inputs from different cone types to signal red versus green (L-M) and blue versus yellow (S-(L+M)). When characterized using detection thresholds, these mechanisms appear to be independent, showing a lack of subthreshold summation (Mullen & Sankeralli, 1998). Above threshold, these pathways interact in a variety of ways, including the enhancement of luminance sensitivity by chromatic contrast, and masking effects between color channels (Chen, Foley, & Brainard, 2000; Cole, Stromeyer, & Kronauer, 1990; DeValois & Switkes, 1983; Mullen, Kim, & Gheiratmand, 2014; Shooner & Mullen, 2020; Switkes et al., 1988). With these interactions not yet fully

understood, it is of interest to search for any differences in how chromatic and achromatic contrast are processed above threshold. In our discrimination experiment we found that dipper functions for all stimuli had a similar form, differing only in the range of physical contrast they occupied. This is consistent with previous studies that showed chromatic and achromatic dippers to overlap when plotted on axes normalized to detection threshold (Switkes et al., 1988). For this type of task, the processing of luminance and color contrast seem to differ only by a scale factor that applies at all contrasts.

For our second experiment (difference-scaling), comparing chromatic to achromatic results requires
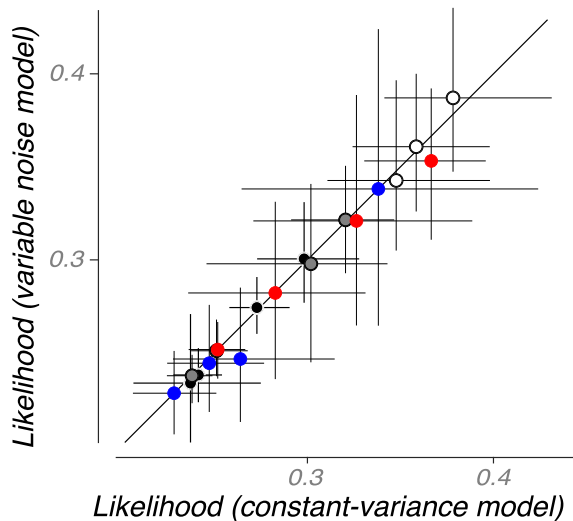
Figure 11. We compared the original MLDS model, with constant-variance noise, to an alternative in which noise variance increased with mean response level. We fit both models to 90% of our data and evaluated the resulting fit's ability to predict the remaining 10% (holdout) data. We repeated this 100 times with different data subsets. Plotted are the median and 95% confidence intervals of the likelihood of the holdout data given the model fit.

consideration of the contrasts chosen for testing. For each condition and observer, we selected a series of contrasts in the range where the discrimination experiment showed a power-law relationship. Each series began above detection threshold and went up in steps of 4 discrimination thresholds. In this sense, the contrast sets were matched perceptually, though in units of absolute contrast they differed both in spacing and range. With test contrast customized in this way, the transducer models resulting from MLDS analysis were similar in form across all stimulus types. When contrast axes were normalized to represent multiples of detection threshold, transducers were highly overlapped for most observers. This allows for a conclusion similar to that of the first experiment: The processing of luminance and color contrast above threshold appear similar, when the very different sensitivity of the separate pathways is taken into account. A similar conclusion applies to the contrast-matching results of Switkes and Crognale (1999), in which observers were asked to compare chromatic to achromatic gratings in terms of apparent contrast. Mapping out the set of contrast pairs that matched perceptually, they revealed a remarkably simple linear relationship in which perceived color and luminance contrast were related by a constant ratio, which was also consistent with ratios of detection thresholds (Switkes, 2008; Switkes & Crognale, 1999).

We also compared stimuli designed to preferentially activate separate achromatic pathways. The spatial and temporal selectivity of magnocellular (M) and parvocellular (P) neurons of the LGN are highly overlapped, and all our achromatic stimuli would be expected to drive responses in both populations. However, as Magnocellular pathways show preference for higher temporal frequencies and lower spatial frequencies, compared with Parvocellular pathways (Lynch, Silveira, Perry, & Merigan, 1992; Merigan, Katz, & Maunsell, 1991), we might expect our high-temporal-frequency flicker stimulus to bias overall neural response toward the M pathway. Similarly, our high-spatial-frequency might be described as "P-biased." These pathways differ substantially in their contrast-response properties: M cells show higher sensitivity to low contrast, but strong saturation at high contrast. P-cells are less sensitive but show a more linear contrast dependence (Benardete & Kaplan, 1999; Kaplan, 2004; Kaplan & Benardete, 2001; Kaplan & Shapley, 1986). To the extent that M cells and their cortical targets contribute to perceived contrast, one might expect our experiments to yield more nonlinear transducer models in the flicker condition. We did not observe this consistently across observers. Similarly, if P cell input dominated in the high spatial frequency condition, we might expect transducers to be more linear in this case. Our results are equivocal on this point. Our discrimination results do not show a stimulus-based difference in the shape of dipper functions that is consistent across observers. Focusing on the high-SF case, the slopes of dipper functions were actually higher than all other conditions in three of four observers, implying a stronger saturation. In the MLDS case, only three observers completed the high-SF condition. In two of these the modeled transducers were in fact more linear (Figures 7 and 8). The small number of subjects does not support a strong conclusion, but motivates further study of the spatial-frequency dependence of apparent contrast as measured with MLDS.

Comparing discrimination and MLDS

The second aim of this study was to compare our two experiments as methods of estimating internal representations of contrast. The two experiments involved very different tasks. Unlike fine discrimination, the difference-scaling task required the observer to make subjective comparisons between stimuli that were clearly different in appearance. Nonetheless, nearly identical models can be used to describe both cases. Both models can be described in terms of an "encoding" stage, where an internal response is assumed to serve as an analog of stimulus contrast, and a "decoding" stage, where responses to multiple stimuli are compared to derive a decision variable. The mean and variance of this variable determine the probability of a given

response via signal-detection methods. A crucial element linking the modeled sensory representation to predicted behavior is an estimate of internal variability, the noise parameter σ in our models.

It is tempting to equate the modeled internal contrast representation with a low-level neural response, such as the summed activity of neurons in V1. We must keep in mind, however, that between the stimulus and the decision-making stage there may be multiple representations of contrast at different levels of the visual hierarchy. The modeled transducer represents the summed contribution of all these stages. This is even more important when interpreting the modeled internal noise, as each stage of processing may contribute to the variability of the final decision variable. We found acceptable agreement in the shape of transducer functions modeled on discrimination and difference-scaling data, but consistently found the scaling experiment to yield higher estimates of internal noise. This result is consistent with the idea that both tasks rely on a common representation of stimulus contrast, but, due to very different task requirements, the scaling experiment is influenced by more levels of uncertainty, up to the level of subjective decision making regarding stimulus appearance.

Several previous studies have compared scaling results to measured discrimination thresholds, but using stimuli and tasks very different from ours. Devinck and Knoblauch (2012) used MLDS to measure perceptual scaling of the watercolor effect. Aguilar et al. (2017) measured perceived tilt from texture. Both studies asked how well MLDS results could predict discrimination thresholds for their respective tasks and found good agreement between the two methods. Our results differ from theirs; the larger internal noise we find in our scaling experiment would lead to a prediction of larger discrimination thresholds than those we measured directly. This difference suggests that MLDS results must be interpreted in the context of the task under study. Lower-level tasks such as contrast discrimination may be influenced by fewer sources of internal variability than more cognitively demanding tasks such as those mentioned above, leading to a greater discrepancy between thresholds and subjective reports on appearance.

Fixed versus variable internal noise

Psychophysical data are most always stochastic: the conditions of interest are those in which the observer gives different responses to the same stimuli on repeated trials, so we measure the probability of a given response. To completely describe a visual process, psychophysical models should account for the source of this variance. Moreover, maximum-likelihood fitting procedures like those used here require models to predict the probability of a given outcome, not just the mean response. Most often this problem is addressed by introducing a random "noise" variable to the model, and most often this noise is assumed to be normally distributed with constant variance. This approach has been successful in many studies of discrimination (Foley, 1994; Georgeson & Meese, 2004; Meese, 2004) and in previous studies using the MLDS approach (Brown et al., 2011; Devinck & Knoblauch, 2012; Knoblauch et al., 2020; Maloney & Yang, 2003). Seemingly at odds with this assumption is the fact that the responses of visual neurons (i.e., spike counts) show variability that increases with mean response (Goris et al., 2014). Variable-noise models can also be used to describe discrimination experiments (Kontsevich et al., 2002); it remains an open question whether fixed or variable noise is most appropriate for describing these results. (Georgeson & Meese, 2006; Kingdom, 2016; Solomon, 2007). The reason for this is that a change in modeled noise can be compensated for by a change in the shape of the model transducer to yield the same predictions. This hinges on the fact that, in any one condition, the two stimuli being discriminated have very similar contrast, and will drive responses with similar variance under either model. Solomon (2007) performed an alternative discrimination experiment designed to isolate internal variability and found support for only a weak dependence of variability on response level. Difference scaling experiments also have the potential to shed light on this problem, as the stimuli within a trial vary more in contrast and would lead to different response variance under a level-dependent noise model.

Single-neuron noise properties provide motivation to test for level-dependent noise psychophysically. However, physiological measurements and computational models suggest that at the level of large populations of neurons, response variability differs in form from that of single units, and shows less dependence on mean response level (Chen, Geisler, & Seidemann, 2006; May & Solomon, 2015). Assuming psychophysical judgements reply on large numbers of neurons, this suggests that there is less of a conflict between constant-variance psychophysical models and neural physiology.

When introducing the MLDS method, Maloney and Yang (2003) tested their assumption of fixed noise using a simulation. They showed that scale values obtained with a fixed-noise model were identical whether the true (simulated) noise was fixed or variable. This is expected given the model structure, where scale values determine the mean of the final decision variable, and modeled noise influences only its variability. They did not detail how the modeled σ behaved in this simulation. Aguilar et al. (2017) performed a similar simulation with a similar result, but also found that agreement between scaling and discrimination results was worse when the

fixed-noise assumption was applied to data generated with variable noise.

We addressed this issue by fitting our data with a modified form of the MLDS model in which variance increased in proportion to mean scale value. Compared to the original fixed-noise model, this alternative was no better or worse at predicted holdout data not used in the fit. However, the parameter of this model that determined the variance-to-mean relationship was consistently found to be quite small, implying a very weak dependence of variance on response level. Although the formal model comparison did not find a winner, the optimization routine appeared to favor the fixed-noise option, on which we based our main findings.

Conclusions

We compared the results of contrast discrimination and contrast difference scaling experiments using a range of simple grating stimuli designed to activate separate visual pathways. Both experiments independently yielded similar estimates of perceived contrast across stimulus conditions, supporting the theory that suprathreshold color and luminance contrast are processed similarly, when the differing sensitivities of these pathways are taken into account (Switkes, 2008; Switkes et al., 1988; Switkes & Crognale, 1999). The results of the two experiments agreed in the shape of the modeled transducer relating physical to perceived contrast, but differed consistently in estimating internal noise, which was higher in the scaling experiment. We conclude that both tasks may rely on a common internal representation of stimulus intensity, but the cognitive demand of the scaling task leads to greater uncertainty than the simpler discrimination task.

Keywords: contrast, color, detection/discrimination, scaling, MLDS

Acknowledgments

Address: McGill Vision Research, Department of Ophthalmology & Visual Sciences, McGill University, Montreal, Quebec, Canada.

References

Aguilar, G., Wichmann, F. A., & Maertens, M. (2017). Comparing sensitivity estimates from MLDS and forced-choice methods in a slant-from-texture experiment. *Journal of Vision, 17*(1), 37–37.

Anstis, S., & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. In: *Colour Vision* (155–166). Toronto: York University.

Benardete, E. A., & Kaplan, E. (1999). Dynamics of primate P retinal ganglion cells: Responses to chromatic and achromatic stimuli. *The Journal of Physiology, 519*(Pt 3), 775–790, http://onlinelibrary.wiley.com/doi/10.1111/j.1469-7793.1999.0775n.x/full.

Bird, C. M., Henning, G. B., & Wichmann, F. A. (2002). Contrast discrimination with sinusoidal gratings of different spatial frequency. *JOSA A, 19*(7), 1267–1273.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Brown, A. M., Lindsey, D. T., & Guckes, K. M. (2011). Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision, 11*(12), 2–2, https://doi.org/10.1167/11.12.2.

Chen, C.-C., Foley, J. M., & Brainard, D. H. (2000). Detection of chromoluminance patterns on chromoluminance pedestals I: Threshold measurements. *Vision Research, 40*(7), 773–788, https://doi.org/10.1016/S0042-6989(99)00227-8.

Chen, Y., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neuroscience, 9*(11), 1412–1420.

Cole, G. R., Hine, T., & McIlhagga, W. (1993). Detection mechanisms in L-, M-, and S-cone contrast space. *Journal of the Optical Society of America A, 10*(1), 38–51, https://doi.org/10.1364/JOSAA.10.000038.

Cole, G. R., Stromeyer, C. F., & Kronauer, R. E. (1990). Visual interactions with luminance and chromatic stimuli. *JOSA A, 7*(1), 128–140, https://doi.org/10.1364/JOSAA.7.000128.

DeValois, K., & Switkes, E. (1983). Simultaneous masking interactions between chromatic and luminance gratings. *JOSA. 73*(1), 11–18, http://www.opticsinfobase.org/abstract.cfm?id=58691.

Devinck, F., & Knoblauch, K. (2012). A common signal detection model accounts for both perception and discrimination of the watercolor effect. *Journal of Vision, 12*(3), 19–19.

Fechner, G. T., Boring, E. G., & Howes, D. H. (1860/1966). *Elements of psychophysics.* Ballwin, MO: Holt, Rinehart and Winston.

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A, 11*(6), 1710, https://doi.org/10.1364/JOSAA.11.001710.

Georgeson, M. A., & Meese, T. S. (2004). Contrast discrimination and pattern masking: Contrast gain control with fixed additive noise. *Movements and Moments in Vision Research. 8th Applied Vision Association Christmas Meeting*, http://eprints.aston.ac.uk/4596/.

Georgeson, M. A., & Meese, T. S. (2006). Fixed or variable noise in contrast discrimination? The jury's still out…. *Vision Research, 46*(25), 4294–4303, https://doi.org/10.1016/j.visres.2005.08.024.

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience, 17*(6), 858–865, https://doi.org/10.1038/nn.3711.

Goris, R. L. T., Putzeys, T., Wagemans, J., & Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychological Review, 120*(3), 472–496, https://doi.org/10.1037/a0033136.

Goris, R. L. T., Wagemans, J., & Wichmann, F. A. (2008). Modelling contrast discrimination data suggest both the pedestal effect and stochastic resonance to be caused by the same mechanism. *JOV, 8*(15), 17.1–21, https://doi.org/10.1167/8.15.17.

Kaplan, E. (2004). The M, P, and K pathways of the primate visual system. *The Visual Neurosciences, 1*, 481–493.

Kaplan, E., & Benardete, E. A. (2001). The dynamics of primate retinal ganglion cells. *Prog Brain Res, 134*, 17–34.

Kaplan, E., & Shapley, R. M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proc Natl Acad Sci USA, 83*(8), 2755–2757.

Kingdom, F. A. (2016). Fixed versus variable internal noise in contrast transduction: The significance of Whittle's data. *Vision Research, 128*, 1–5.

Klein, S. A. (2006). Separating transducer non-linearities and multiplicative noise in contrast discrimination. *Vision Research, 46*(25), 4279–4293, https://doi.org/10.1016/j.visres.2006.03.032.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?

Knoblauch, K., & Maloney, L. T. (2008). MLDS: Maximum likelihood difference scaling in R. *Journal of Statistical Software, 25*(2), 1–26.

Knoblauch, K., Marsh-Armstrong, B., & Werner, J. S. (2020). Suprathreshold contrast response in normal and anomalous trichromats. *JOSA A, 37*(4), A133–A144, https://doi.org/10.1364/JOSAA.380088.

Kontsevich, L. L., Chen, C.-C., & Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research, 42*(14), 1771–1784, https://doi.org/10.1016/S0042-6989(02)00091-3.

Kulikowski, J. J. (1976). Effective contrast constancy and linearity of contrast sensation. *Vision Research, 16*(12), 1419–1431.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70*(12), 1458–1471, https://doi.org/10.1364/JOSA.70.001458.

Lynch, J. J., Silveira, L. C. L., Perry, V. H., & Merigan, W. H. (1992). Visual effects of damage to P ganglion cells in macaques. *Visual Neuroscience, 8*(06), 575–583, https://doi.org/10.1017/S0952523800005678.

Maloney, L. T., & Knoblauch, K. (2020). Measuring and Modeling Visual Appearance. *Annual Review of Vision Science, 6*, 519–537.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision, 3*(8), 5–5.

May, K. A., & Solomon, J. A. (2015). Connecting psychophysical performance to neuronal response properties I: Discrimination of suprathreshold stimuli. *Journal of Vision, 15*(6), 8–8.

Meese, T. S. (2004). Area summation and masking. *Journal of Vision, 4*(10), 930–943, https://doi.org/10.1167/4.10.8.

Merigan, W., Katz, L., & Maunsell, J. (1991). The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys. *The Journal of Neuroscience, 11*(4), 994–1001.

Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology, 359*(1), 381–400, https://doi.org/10.1113/jphysiol.1985.sp015591.

Mullen, K. T., Kim, Y. J., & Gheiratmand, M. (2014). Contrast normalization in colour vision: The effect of luminance contrast on colour contrast detection. *Scientific Reports, 4*, 7350, https://doi.org/10.1038/srep07350.

Mullen, K. T., & Losada, M. A. (1994). Evidence for separate pathways for color and luminance detection mechanisms. *JOSA A, 11*(12), 3136–3151, https://doi.org/10.1364/JOSAA.11.003136.

Mullen, K. T., & Sankeralli, M. J. (1998). Evidence for the stochastic independence of the blue-yellow, red-green and luminance detection mechanisms revealed by subthreshold summation. *Vision Research, 39*(4), 733–745, https://doi.org/10.1016/S0042-6989(98)00137-0.

Obein, G., Knoblauch, K., & Viéot, F. (2004). Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *Journal of Vision, 4*(9), 4, https://doi.org/10.1167/4.9.4.

Olzak, L. A., & Thomas, J. P. (2003). Dual nonlinearities regulate contrast sensitivity in pattern discrimination tasks. *Vision Research, 43*(13), 1433–1442.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *JOSA A, 2*(9), 1508–1532.

Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Sankeralli, M. J., & Mullen, K. T. (1996). Estimation of the L-, M-, and S-cone weights of the postreceptoral detection mechanisms. *JOSA A, 13*(5), 906–915, https://doi.org/10.1364/JOSAA.13.000906.

Sankeralli, M. J., & Mullen, K. T. (1997). Postreceptoral chromatic detection mechanisms revealed by noise masking in three-dimensional cone contrast space. *JOSA A, 14*(10), 2633–2646, https://doi.org/10.1364/JOSAA.14.002633.

Shooner, C., & Mullen, K. T. (2020). Enhanced luminance sensitivity on color and luminance pedestals: Threshold measurements and a model of parvocellular luminance processing. *Journal of Vision, 20*(6), 12–12, https://doi.org/10.1167/jov.20.6.12.

Smith, V. C., & Pokorny, J. (1975). Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research, 15*(2), 161–171, https://doi.org/10.1016/0042-6989(75)90203-5.

Solomon, J. A. (2007). Contrast discrimination: Second responses reveal the relationship between the mean and variance of visual signals. *Vision Research, 47*(26), 3247–3258, https://doi.org/10.1016/j.visres.2007.09.006.

Switkes, E. (2008). Contrast salience across three-dimensional chromoluminance space. *Vision Research, 48*(17), 1812–1819.

Switkes, E., Bradley, A., & DeValois, K. K. D. (1988). Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings. *J Opt Soc Am A, 5*(7), 1149–1162.

Switkes, E., & Crognale, M. A. (1999). Comparison of color and luminance contrast: Apples versus oranges? *Vision Research, 39*(10), 1823–1831, https://doi.org/10.1016/S0042-6989(98)00219-3.

Whittle, P. (1992). Brightness, discriminability and the "Crispening Effect". *Vision Research, 32*(8), 1493–1507, https://doi.org/10.1016/0042-6989(92)90205-W.