Contributions of contour frequency, amplitude, and luminance to the watercolor effect estimated by conjoint measurement

Peggy Gerardin
INSERM U846 Stem-Cell and Brain Research Institute, Department of Integrative Neurosciences, Bron, France
Université Lyon 1, Lyon, France

Frédéric Devinck
Université Rennes 2, CRPCC, Rennes, France

Michel Dojat
INSERM U836, Université Grenoble Alpes, GIN, Grenoble, France

Kenneth Knoblauch
INSERM U846 Stem-Cell and Brain Research Institute, Department of Integrative Neurosciences, Bron, France
Université Lyon 1, Lyon, France

Introduction

The watercolor effect (WCE), first described by Pinna (1987), has provoked great interest over the last decade (Pinna, Brelstaff, & Spillmann, 2001). In its classic configuration, it is generated by two adjacent and thin, parallel, wavy contours of different chromaticity separating two regions in the visual field. The color of one of the contours appears to diffuse into the region it borders, filling it with a washed-out shade of the contour. The fill-in can occur over a large region of the visual field, raising the possibility that the phenomenon relates to high-level processes (Spillmann & Werner, 1996), such as surface perception (Pinna et al., 2001), figure-ground segregation (Pinna, Werner, & Spillmann, 2003; von der Heydt & Pierson, 2006; Pinna & Tanca, 2008; Tanca & Pinna, 2008), or other aspects of perceptual organization (Pinna, 2005).

The strength of the WCE is influenced by several factors, including the relative luminances of the contours (Devinck, Delahunt, Hardy, Spillmann, & Werner, 2006; Devinck & Knoblauch, 2012), the width of the inducing contours (Pinna et al., 2001; Devinck, Gerardin, Dojat, & Knoblauch, 2014), and the continuity and contiguity of the contour pairs (Devinck & Spillmann, 2009). Most studies have assessed the strength of the effect with matching or hue cancellation...
techniques (Devinck, Delahunt, Hardy, Spillmann, & Werner, 2005; von der Heydt & Pierson, 2006). More recently, paired comparisons have been used to quantify the phenomenon. Cao, Yazdanbakhsh, and Mingolla (2011) judged which of a pair of stimuli showed a more salient brightness filling-in for an achromatic version of the WCE. The strength of the WCE was related to the probability of distinguishing two patterns. Devinck and Knoblauch used maximum likelihood difference scaling (MLDS) to estimate the strength of the WCE as a function of the luminance of the interior inducing contour. MLDS involves comparisons between large stimulus differences or intervals and leads to an interval scale that describes changes in stimulus appearance along a single dimension (Maloney & Yang, 2003; Knoblauch & Maloney, 2012b).

MLDS is a promising technique because it is based on a Gaussian, equal-variance signal-detection model (Knoblauch & Maloney, 2012b). The estimated perceptual scales have been shown to predict discrimination behavior, thus yielding an integrated description of performance from threshold to perception (Devinck & Knoblauch, 2012). A disadvantage is that the method only permits evaluation of one stimulus dimension at a time. This difficulty, however, is overcome by the paradigm of conjoint measurement (Luce & Tukey, 1964; Krantz, Luce, Suppes, & Tversky, 1971; Falmagne, 1985; Roberts, 1985; Knoblauch & Maloney, 2012b), which permits estimating interval scales for two or more dimensions from the same experimental data. In this paradigm, the observer judges one pair of stimuli in each trial, but the levels of the dimensions being studied vary independently across the two stimuli from trial to trial. In this way, the influence on judgments of varying one dimension while the other is fixed can be evaluated. Ho et al. (2008) recently developed a Gaussian, equal-variance, signal-detection model of conjoint measurement and used maximum likelihood to estimate the underlying perceptual scales and to test several hypotheses on how observers combine information across dimensions (maximum likelihood conjoint measurement or MLCM). We used this procedure to study the joint influences of luminance contrast, contour frequency, and amplitude on the WCE.

Material and methods

Observers

Six observers (two male and four female) with normal or corrected-to-normal vision volunteered for the experiments (mean age ± SD: 33 ± 8 years). Three participated in all three conditions, two in only one condition, and one in two of the conditions. All observers but one (author PG) were naive, and all had normal color vision as assessed by a Farnsworth Panel D15. Observers who required optical corrections wore their glasses while performing the experiments. All experiments were in accordance with the principles of the Declaration of Helsinki.

Apparatus

The experiments were performed in a dark room. Stimuli were displayed on an Eizo FlexScan T562-T color monitor (42 cm) driven by a MacBook Pro (2.2 GHz). The screen had a resolution of 800 × 600 pixels and was run at a field rate of 100 Hz, noninterlaced. The voltage-phosphor luminance relationship was linearized with look-up tables. Calibration of the screen was performed with a Minolta CS-100 chroma meter. Observers were placed at a distance of 57.3 cm from the screen.

Stimuli

Each stimulus was generated with Matlab R2009b (http://www.mathworks.com/) and displayed with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). All stimuli were displayed on a white background (128 cd/m², CIE xy = 0.29, 0.30). The contour pairs that defined the stimuli were each of width 16 min, that is, 8 min for the interior and exterior contours, each. The outer contour was purple (CIE xy = 0.32, 0.19) and the inner orange (CIE xy = 0.48, 0.34). The stimuli were also specified in the DKL color space (Derrington, Krauskopf, & Lennie, 1984) with purple and orange contours at azimuths of 320° and 45°, respectively. Control stimuli were identical except that the contours were interlaced and generated little filling-in (Figure 1c).

For independent control of the frequency and the amplitude of the WCE contours, they were constructed as Fourier descriptors (Zahn & Roskies, 1972). Each stimulus was defined by a circle of 4° diameter whose radius was modulated sinusoidally as a function of angle according to the equation

\[ R(\theta) = r + A \sin(2\pi f \theta), \]

where \( R \) is the stimulus radius at angle \( \theta \), \( r \) the average radius of the stimulus, \( A \) the modulation, and \( f \) the frequency in cycles per revolution (cpr). The stimuli are then plotted by transforming the polar coordinates \( (R, \theta) \) to rectangular coordinates \( (x, y) \) with the equations

\[ x(R, \theta) = R(\theta) \sin(2\pi \theta) \]  
\[ y(R, \theta) = R(\theta) \cos(2\pi \theta). \]
For each condition, a factorial design was defined in which five levels along each of two dimensions (frequency, amplitude, or luminance) were chosen, and all levels crossed, creating a table of 25 stimuli. Figure 1a shows an example for frequency and amplitude with frequency varying across columns and amplitude across rows. The five frequencies were equally spaced, ranging from 4 to 20 cpr. The five amplitudes evaluated ranged from 0.04 to 0.2, and the five luminances of the orange contour ranged from 0.1 to 0.9 equally spaced in elevation in DKL space. In two other conditions, the dimensions of frequency/luminance and amplitude/luminance were similarly crossed.

Procedure

Observers were run in three conditions (luminance/frequency, luminance/amplitude, frequency/amplitude) with a single condition tested within a session. In each trial, a pair of different stimuli from the stimulus table for that condition (e.g., Figure 1a) was chosen at random and presented to the observer. Observers judged which of the two patterns evoked the most salient fill-in color. To test that observers responded according to the fill-in appearance and not on the basis of some other stimulus feature(s), an equal number of test and control stimuli were interleaved in each session (e.g., Figure 1b and c). There are $25 \times 24/2 = 300$ such pairs. In each trial, either a pair of test or control stimuli were presented. A session consisted of a random presentation of all 600 test and control pairs. Each condition was repeated in five sessions, yielding 1,500 test and 1,500 control trials per condition for a given observer.

Model

We modeled the data following the framework described by Ho et al. (2008) for analyzing conjoint measurement experiments (see also chapter 8, Knoblauch & Maloney, 2012b). They define three nested models that can be fit to describe the choices of the observers in order of complexity: an independence model, an additive model, and a saturated model. We begin the description with the additive model and then describe the simpler and more complex models with respect to it. We represent the stimulus levels along the two dimensions of a condition by a variable $\phi_{ij}$, $i, j = 1, \ldots 5$, where the indices refer to two of the dimensions in
the set (luminance, frequency, amplitude), for example, a row and column, respectively, of Figure 1a. In the additive model, we suppose that each of the two dimensions contributes to a filling-in response, $\psi_{ij}$, respectively, which depends on its physical intensity level, and that the filling-in response, $\psi_{ij}$, to a stimulus, $\phi_{ij}$, is the sum of the component responses

$$\psi_{ij} = \psi_{ij}^1 + \psi_{ij}^2.$$  

(4)

We assume that the observer chooses the first stimulus exactly when the decision rule is equivalent to a comparison of intervals across dimensions.

It is unlikely, however, that the observer would make the same response in every trial when the two stimuli are very similar. To incorporate this inherent variability of human responses, we suppose that the decision variable, $\Delta_{ijkl}$, is contaminated by internal noise so that the observer chooses the first stimulus exactly when

$$\Delta_{ijkl} = \psi_{ijkl}^1 - \psi_{ijkl}^2 + \epsilon_{ijkl} > 0,$$

(7)

where each $\epsilon_{ijkl}$ is a draw from a distribution of independent and identically distributed normal variables with $\mu = 0$ and variance $= 4\sigma^2$. This is an equal-variance, Gaussian signal-detection model. The coefficient of four on the variance parameterizes the estimated scale values so that variance of the response along each dimension is equal to $\sigma^2$. As a result, the estimated response values $\psi_{ijkl}/\sigma$ are distributed as normal variables with $\sigma^2 = 1$ and are, thus, on the same scale as the sensitivity measure $d'$ from signal detection theory (Green & Swets, 1966).

With $p$ intensity levels sampled along each dimension and the estimate of the variance of $\epsilon$, there are $2p + 1$ parameters in the model. The reparameterizations of the scale described above suggest that we may multiply the estimated values by any constant without changing the model predictions of the observer’s responses. In fact, we may similarly add any constant to the scales with similar results. To fix the estimated scales, we set the lowest value of each scale to zero.

This eliminates two parameters from the fit. In addition, by making $\sigma^2 = 1$, we eliminate one more so that the fitting process requires estimating $2p - 2$ parameters. The parameterization that we propose is slightly different from that used by Ho et al. (2008), who estimated $\sigma$ but normalized the estimated functions so that the maximum scale value was unity.

Both parameterizations yield identical predictions of performance.

The additive model implies that there are component functions representing the internal response to each physical dimension and that the overall response to a stimulus is the simple sum of these component responses. We illustrate this in Figure 2c and d. In Figure 2c, we show two hypothetical component responses, $\psi^1$ and $\psi^2$, each of which, for simplicity, is a linear function of the stimulus level. Stimulus level here and elsewhere is indicated by an index, not physical units. This convention, as used elsewhere (Ho et al., 2008; Knoblauch & Maloney, 2012b), allows the scales for both dimensions to be plotted together. The response to any stimulus with levels $i$ and $j$, respectively, along the two stimulus dimensions is represented by the sum of their component responses. As an example, consider a stimulus of level three along the first dimension and level four along the second. The component responses to the level along each dimension are indicated by the two black points. Figure 2d shows the set of summed responses for all pairings of the two dimensions with the base dimension corresponding to $\psi^2$ and the parameter indicated along each curve corresponding to $\psi^1$. Each curve has the shape of the component curve for $\psi^2$ but is displaced vertically by the value of $\psi^1$, resulting in a set of parallel contours. Analogous to a linear model, we could say that each dimension shows a main effect but no interaction. The summed response to the stimulus with the levels, indicated in Figure 2c is indicated by the black point.

A simpler model than the additive model is obtained by assuming that the observer’s judgments depend on only one of the component dimensions. In that case, the decision variable reduces to

$$\Delta_{ijkl} = \psi_{ijkl}^1 - \psi_{ijkl}^2 + \epsilon_{ijkl}$$

(8)

where we assume that only the first dimension contributes to the judgments. Ho et al. (2008) call this the independent-property model, but we will refer to it as just the independence model. The independence model requires estimating only $p - 1$ parameters.

The independence model is illustrated in Figure 2a and b. In Figure 2a, the component $\psi^1$ increases with stimulus level along the first dimension, but $\psi^2$ is flat or independent of the level of the second dimension. The summed responses are indicated in Figure 2b by a set of lines of zero slope that are vertically displaced by the responses along the $\psi^1$ curve. In analogy to a linear model, there is a significant main effect of the first dimension but not of the second.

Finally, the additive model may not suffice to provide a satisfactory description of the observer’s responses. Ho et al. (2008) describe the full model, which includes interaction terms, $\psi_{ijkl}^1$, that depend on the intensity levels of both dimensions. The decision
variable is
\[ \Delta_{ijkl} = (\psi_1^i + \psi_2^j + \psi_{12}^{ij}) - (\psi_1^k + \psi_2^l + \psi_{12}^{kl}) + \epsilon_{ijkl}. \]

This model requires the estimation of \( p^2 - 1 \) parameters, one less than the number of stimuli tested, and will be referred to as the saturated model. An example of response curves that might result from a saturated model are shown in Figure 2e. The important feature is that the curves are not parallel and, thus, the curves cannot be explained by a simple additive combination of component curves as in the previous two cases.

**Results**

Judgments from the observers for test and control conditions for each of the pairings of the three dimensions are shown in Figure 3a through c in a format introduced by Ho et al. (2008) and termed a **conjoint proportion plot** (CPP) by Knoblauch and Maloney (2012a). Each CPP summarizes the proportion of times that the ordinate stimulus, \( S_{kl} \), was judged to show a greater fill-in than the abscissa stimulus, \( S_{ij} \), coded according to the grey levels shown in the color bar at the right of each set of graphs, for every stimulus pair presented. The levels along the two dimensions are...
represented along each axis using a scheme in which a 5 × 5 outer grid demarcates the stimulus levels along one dimension (e.g., amplitude in Figure 3a), and the levels of the second dimension are represented as an inner subgrid nested within each square of the outer grid (e.g., frequency in the same figure). As the responses are combined across random left/right orders in the presentation—only the upper left triangle of a CPP is unique and displayed.

In Figure 3a through c, the set of results from each observer is indicated by his/her initials above the pair of CPPs for test (upper) and control (lower) stimuli. Figure 3d shows the expected pattern of responses for an ideal observer for whom the judgments are based on the level of only one of the two stimulus dimensions. The left CPP shows the results when the judgments depend solely on the outer dimensions (Dim 1) and the right on the inner (Dim 2). In general, the CPP for the test stimuli resemble more closely the ideal pattern displayed for the inner dimension, suggesting that the contribution of frequency dominated when both frequency and amplitude covaried but that luminance contributed more strongly to the judgments when paired with the other two dimensions. The patterns do, however, deviate from the ideal, indicating contributions from both dimensions.

There is an important difference in the procedure used here from that reported by Ho et al. (2008). They tested stimuli that varied along the dimensions of roughness and glossiness and asked different observers to judge which member of a pair appeared rougher or shinier. Thereby, they could estimate the contamination of the one dimension to judgments along the other. In our experiments, the task is not necessarily related to the stimulus dimensions. Thus, one cannot be sure a priori that the observer is actually basing his/her judgments on the stated criterion. Instead, the observer may simply be judging the variation along the two dimensions manipulated, that is, which stimulus has a higher frequency contour, amplitude, or luminance. If observers were performing in this manner, then we would expect to see the same pattern of responses in the CPPs for control and test stimuli. By and large, however, the patterns are quite different with the control CPP patterns appearing to be random. The principal exception to this is evident in the control data.

Figure 3. Conjoint proportion plots. Each plot shows the proportion of stimulus, $S_{ij}$, judged to show a greater fill-in than the stimulus represented on the abscissa, $S_{k}$, as grey level coded according to the color bar at the right. Levels $i$ and $k$ along one dimension are indicated by the exterior axis numbers. Each of the squares $(i, k)$ is subdivided in a 5 × 5 grid indicating the pairings $(j, l)$ along the second dimension. Interior axis labels are shown for the extreme levels in the two lower right squares of each plot. In each set of plots, the top row indicates the results for the test stimuli and the bottom for the control with observers’ initials indicated in the upper strips. (a) Outer/inner dimensions: amplitude/frequency. (b) Amplitude/luminance. (c) Frequency/luminance. (d) Expected response patterns for an observer who judges the stimuli based on the contributions along only one of the dimensions. Dim 1 shows the expected pattern for the outer dimension and Dim 2 for the inner dimension.
of observer OH for the amplitude/luminance condition (Figure 3b) and, to a lesser extent, for observer DL in the same condition. In these two cases, then, we cannot exclude that either the observers perceived a small amount of filling-in or, alternatively, that they showed a tendency to base their judgments on the physical dimensions of the stimulus for a subset of the trials rather than on the perceived filling-in per se.

Figure 4a through c shows the average estimated scales under the additive model for each observer for each pairing of frequency, luminance, and amplitude. A different color is used to indicate the contributions to the judgment of each dimension: red = frequency, blue = amplitude, green = luminance. The top row of graphs in each subfigure shows the scale values estimated for the test stimulus and the bottom row for the control.

For example, Figure 4a indicates that the contribution of the frequency dimension to the filling-in strength increases with the frequency of the contour while the amplitude dimension shows a smaller effect that increases between the first and second levels tested for all observers but one (SJ shows a relatively flat amplitude contribution) and then remains approximately constant. These results contrast with the scales estimated for the control stimuli that show no difference and very little effect of either dimension on the strength of filling-in.

Similarly, Figure 4b shows that the contribution of luminance to the filling-in increases with the luminance of the orange contour while the amplitude contributions are similar to those obtained in Figure 4a with the exception of observer SJ who shows a stronger
luminance contribution than in the previous condition. The controls show little influence of either dimension. The exception is again shown by observer OH, consistent with the CPP plots in Figure 3b. Note also, that, for this condition, the responses of observers PG and OH to the test stimuli result mostly from the luminance dimension, again, as suggested by comparison of their CPP plots to the ideal cases.

Figure 4c shows the contributions of luminance and frequency when these two dimensions are varied. The luminance scales are similar to those in Figure 4b, and the frequency scales resemble those obtained in Figure 4a. As in the preceding figures, the control stimuli provide little evidence for filling-in that the contributions of each dimension to the judgments remain low as the scale value is increased.

Finally, Figure 4d shows the means and 95% confidence intervals across observers for each pairing of the three dimensions. The scales for the control conditions show little effect with confidence intervals overlapping zero, and confidence intervals for test conditions exclude zero with the exception of the highest amplitude in the amplitude/luminance condition. The large confidence intervals for some of the test conditions reflect the individual differences shown in the previous graphs. Interestingly, the average scale for each dimension tends to be similar, independently of the other dimension with which it is paired, a point we will examine in more detail below.

Comparisons of independence and additive models

We consider the possibility that the contribution of only one dimension independently of the other suffices to account for observers’ judgments, that is, the independence model. We chose to model the data based on the dimension that generated the strongest contributions under the additive model. We will refer to this dimension as the primary dimension and the other as the secondary dimension. Specifically, luminance was the primary dimension when paired with frequency or amplitude, and frequency was the primary dimension when paired with amplitude.

We compare the models by plotting the estimated scales from both models in the same graph for each observer. For reference, we are comparing the two models shown in Figure 2a and c. Comparisons between additive (solid) and independence (grey dashed) model fits are displayed in Figure 5 with 95% confidence intervals of the additive estimates displayed. By construction, the contribution of the secondary dimension is fixed at zero for the independence model.

The comparisons in Figure 5 indicate that scale estimates of the primary dimension under the inde-
for the contribution of luminance depend on whether either frequency or amplitude is the covarying dimension? The same question can be posed for each dimension with respect to the other two. Figure 8a shows the estimated contribution of luminance to the fill-in color when the contour frequency (red) or its amplitude (blue) is the covarying dimension as a function of the luminance index for the four observers who performed both experiments. The scales obtained under the different conditions agree very well for all of the observers, and the overlap in the confidence intervals lends no support to the importance of the differences in the means.

Similarly, Figure 8b and c shows the comparisons of the scales obtained for the frequency and amplitude dimensions, respectively. There is good agreement across conditions for the frequency scales for observers PG and SJ. For OH, the scales agree in form although the sensitivity to frequency is higher when paired with the luminance than with the amplitude dimension.

There seems to be less agreement between the scales based on the amplitude dimension. This is most marked for observer SJ, for whom the contribution of amplitude to the judgments is insignificant when paired with frequency but significantly higher, based on the nonoverlapping of confidence intervals, when paired with the luminance dimension. As noted

---

Figure 5. Comparisons of additive and independent model fits for each observer in the three experiments. The solid lines indicate the estimated contributions of each dimension under the additive model, and the dashed grey lines indicate the estimates under the independence model. The color codes correspond to those of the points from Figure 4 with red = frequency, green = luminance, and blue = amplitude. The error bars are bootstrap estimated 95% confidence intervals (n = 10,000) under the additive model.
(a) Amplitude/frequency. (b) Amplitude/luminance. (c) Frequency/luminance.
previously (see Figure 4a and b), however, the amplitude contributions tended to be small and more variable. This is borne out by the averages across observers shown in Figure 8d that show good agreement in the scales measured along each dimension and agrees with the averages based on all observers from Figure 4d. The average data show the amplitude contribution asymptotes by the second stimulus level, which corresponds to an amplitude of 0.08. The frequency contribution asymptotes by the third level, which corresponds to a value of 12 cpr.

**Discussion**

We used the MLCM technique to quantify the contributions to the filling-in strength of the WCE of three stimulus dimensions. An important difference from the previous use of this technique by Ho et al. (2008) is that the perceptual judgment required of the observers in our experiments is not directly linked to either of the stimulus dimensions whereas in Ho et al. (2008) it was. Thus, in theory, the observers could have
simply based their judgments directly on the stimulus dimensions manipulated (luminance, frequency, and/or amplitude) rather than the perceptual criterion of filling-in strength. To evaluate this possibility, we included control stimuli during the sessions for which very little, if any, filling-in was expected to occur. If the observers based their judgments on the manipulated dimensions rather than the perceived filling-in, then the estimated contributions to the judgments for control and test stimuli would be the same. In fact, the model estimates for the control conditions were severely attenuated in all conditions, thus supporting that the test conditions do reflect judgments with respect to filling-in.

As shown previously using a difference scaling technique (Devinck & Knoblauch, 2012), the strength of the WCE depends importantly on the luminance of the interior contour. We show here, as well, that the filling-in is stronger for higher frequencies of the contour modulation although the influence of contour frequency asymptotes at about 12 cpr. Probably because of its small value, the estimates of the contribution of amplitude to the phenomenon were found to be more variable across conditions and observers. In addition, the results are best supported by a model in which the responses to the separate dimensions contribute in an additive fashion.

It has previously been reported that the WCE is present although weaker with straight contours (Pinna et al., 2001). Our results confirm this observation in that, for nearly all conditions, the amplitude contribution was greater for higher amplitudes than for the lowest value tested. We did not test a frequency of zero cpr or a zero amplitude because it would have led to a singular set of conditions, a row and column of identical stimuli in the frequency/amplitude matrix of

Figure 7. Residual differences between additive and saturated model fits shown in Figure 6 for all observers and conditions plotted as grey levels according to color bars at the right of each plot. (a) Residuals for model fits of frequency/amplitude pairing for each observer. (b) Residuals for models fits of amplitude/luminance pairing. (c) Residuals for models fits of frequency/luminance pairing. The residuals have been normalized with respect to the estimated standard errors for the saturated model fits.
stimuli (Figure 1). Additionally, because we fix the lowest perceptual scale values at zero in order to make the estimated coefficients identifiable, our measurements of the strength of the phenomenon are only in reference to this value. A measure of the strength at this lowest value could have been obtained by including paired comparisons between test and control stimuli in the experiment, but this would have quadrupled the number of stimulus pairs to evaluate.

We performed the experiments analyzing dimensions pairwise, that is, two-way experiments. Following the logic of factorial design, it might be argued that it would be more efficient to perform an experiment with all three attributes permitted to covary simultaneously, that is, a three-way experiment. Employing five levels along each dimension, as here, would generate $5^3 = 125$ different stimuli and $125 \times 124/2 = 7,750$ paired comparisons, which could be reasonably argued is excessive. Simulation results for two-way experiments suggest that it is the number of total judgments and not the number of conditions that determines the precision of the estimates (Knoblauch & Maloney, 2012b). Similar results have been obtained for the MLDS technique (Maloney & Yang, 2003). This raises the possibility of obtaining good scale estimates for a three-way experiment by subsampling from the full set of stimulus pairs although it will be necessary to verify such a conjecture via simulation.

Comparing the dimensions pairwise permitted us to evaluate whether the estimated contributions of each dimension to the observer’s judgments were independent of the contributions of the second, covarying dimension. The results support such an invariance for the luminance and frequency dimensions but are more equivocal for the amplitude although all three dimensions are consistent with invariance for the average data. We suspect that the failures of invariance along the amplitude dimension stem more from its weak contribution to the WCE, rendering it difficult to estimate with the number of trials that we employed and possible criterion shifts of the observers. These hypotheses remain to be tested, however.

Figure 8. (a) Comparison of the average perceptual scales as a function of contour luminance for four observers. Red symbols correspond to the experiment frequency $\times$ luminance and blue symbols to the experiment amplitude $\times$ luminance. Each point is the mean of five sessions, and the error bars are 95% confidence intervals. (b) Comparison of the average perceptual scales as a function of contour frequency for three observers. Green symbols correspond to the experiment frequency $\times$ luminance and blue symbols to the experiment amplitude $\times$ frequency. Each point is the mean of five sessions, and the error bars are 95% confidence intervals. (c) Comparison of the average perceptual scales as a function of contour amplitude for three observers. Red symbols correspond to the experiment amplitude $\times$ luminance and green symbols to the experiment amplitude $\times$ frequency. Each point is the mean of five sessions, and the error bars are 95% confidence intervals. (d) Mean comparison of the average perceptual scales as a function of the indices of contour amplitude (left, $n = 3$), of contour frequency (middle, $n = 3$), and of contour luminance (right, $n = 4$). Blue symbols correspond to experiments with amplitude, green symbols to experiments with luminance, and red symbols to experiments with frequency. The error bars are 95% confidence intervals based on interindividual differences.
The WCE depends on both local spatial and chromatic properties of the contour of the stimulus but generates a long-range filling-in percept as well as a figure-ground segregation between the fill-in region and the surround. Models to account for the WCE, therefore, typically include multiple levels of processing (Pinna et al., 2001; Pinna & Grossberg, 2005; von der Heydt & Pierson, 2006). We have recently shown (Devinck et al., 2014), for example, that the dependence of the strength of the WCE on the spatial frequency of the inducing contour width has a narrow tuning similar to a class of luminance-chromatic cells found in cortical area V1 of the macaque (Shapley & Hawken, 2011). Nevertheless, the spatial extent of V1 receptive fields and their interactions would seem to preclude this site as the substrate for the long-range filling-in phenomenon. Sensitivity to the frequency of the contours also seems unlikely to be a property of area V1 because selectivity for contour curvature is only observed in higher-order visual areas (Schwartz, Desimone, Albright, & Gross, 1983; Hegdé & Van Essen, 2000; Roe et al., 2012) and modeled as such (Habak, Wilkinson, Zakher, & Wilson, 2004; Poirier & Wilson, 2006). Our results do not distinguish directly between a mechanism broadly tuned to the contour frequency and an ensemble response over individual elements that are narrowly tuned. The fact that the additive model adequately explains the results, however, can be taken as consistent with a mechanism whose response is invariant for a trade-off of the response along one dimension with respect to that along another; that is, the response of increase along one dimension could be compensated by decrease along the other. Such an interpretation would favor a single mechanism with broad tuning to curvature.

Keywords: psychophysics, scaling, conjoint measurement, filling-in, color vision, assimilation, watercolor effect

Acknowledgments

This work was supported by a grant, ANR 11JSH20021, from the Agence Nationale de la Recherche to Frédéric Devinck. John S. Werner is gratefuly thanked for comments.

Commercial relationships: none.
Corresponding author: Kenneth Knoblauch. Email: ken.knoblauch@inserm.fr.
Address: INSERM U846 Stem-Cell and Brain Research Institute, Department of Integrative Neurosciences, Bron, France and Université Lyon 1, Lyon, France.

Footnote

1See Devinck and Knoblauch (2012) and Knoblauch and Maloney (2012b), p. 202, for the derivation for MLDS and Devinck and Knoblauch (2012) for empirical verification. The MLCM model is formally the same as that for MLDS except for two sign changes in the decision variable, and therefore, the same ideas should apply.

References


