Prior implicit knowledge shapes human threshold for orientation noise

Jeppe H. Christensen
Department of Psychology, University of Copenhagen, Copenhagen, Denmark

Peter J. Bex
Department of Psychology, Northeastern University, Boston, MA, USA

József Fiser
Department of Cognitive Science, Central European University, Budapest, Hungary

Although orientation coding in the human visual system has been researched with simple stimuli, little is known about how orientation information is represented while viewing complex images. We show that, similar to findings with simple Gabor textures, the visual system involuntarily discounts orientation noise in a wide range of natural images, and that this discounting produces a dipper function in the sensitivity to orientation noise, with best sensitivity at intermediate levels of pedestal noise. However, the level of this discounting depends on the complexity and familiarity of the input image, resulting in an image-class-specific threshold that changes the shape and position of the dipper function according to image class. These findings do not fit a filter-based feed-forward view of orientation coding, but can be explained by a process that utilizes an experience-based perceptual prior of the expected local orientations and their noise. Thus, the visual system encodes orientation in a dynamic context by continuously combining sensory information with expectations derived from earlier experiences.

Introduction

Accurate representation of local contour orientation in the visual system is thought to be crucial for everyday object and scene perception (Biederman, 1987; Marr, 1982). Driven by this notion, the neuro-physiological (Hubel & Wiesel, 1968; Somers, Nelson, & Sur, 1995) and psychophysical (Graham, 1989) bases of both detecting and encoding orientation information have been widely investigated, and the results have become fundamental components of present-day theories of vision. According to the prevalent view, spatial vision in both humans and animals is based on the outputs of local sensors that encode the orientation of small regions of visual space (Campbell & Kulikowski, 1966). Therefore, traditional investigations of orientation coding in humans have been conducted with lines or Gabor-patches in simple orientation discrimination tasks (Field & Tolhurst, 1986). Setting the stage for computational models of visual processing (Fukushima, 1980; Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007), the assumption based on these studies has been that orientation information at this fixed, high precision is used and accessible during the perception of more complex visual forms and shapes. However, it is not clear how these results generalize to everyday processing of orientation information in natural scenes, where stimuli are typically of much higher complexity and must be recognized across many different viewing angles and projections. Indeed, both psychophysical (Adelson, 1993; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) and physiological (Ringach, Hawken, & Shapley, 1997) data show that for more complex stimuli, human behavioral as well as neural responses to various visual attributes are strikingly different from those measured with simple stimuli. Furthermore, typical effects found under threshold conditions seem not to hold under suprathreshold conditions (Fiser, Bex, & Makous, 2003). Thus, in order to properly assess the precision of human orientation coding during everyday vision and under suprathreshold conditions, a systematic quantitative measure of orientation sensitivity is needed with
stimuli that more closely approximate the complexity of natural visual inputs.

Measuring the precision of orientation coding under natural conditions is challenging, as traditional methods of assessing the sensitivity to a single well-isolated oriented visual element do not properly convey the level of coding in naturalistic vision (Olshausen & Field, 1996). To handle this issue, we developed a novel image-manipulation technique to assess how the human visual system represents the orientation structure of a wide variety of visual stimuli (Figure 1A). In our method, images are decomposed into their local representation by a bank of wavelet (Gabor) filters (Daugman, 1985) similar to those found in the primary visual cortex of mammalian brains (DeValois & DeValois, 1988). This process can then be reversed to generate a stimulus composed of a subset of the obtained Gabor elements with known position, contrast, phase, and orientation that can reconstruct the original image with any specified precision. To obtain a global assessment of the precision of orientation coding with naturalistic stimuli, we measured human sensitivity to changes in standard deviation (SD) of local contour orientation while varying the parameters of the Gabor-patches that constitute the stimulus. Specifically, we generated classes of stimuli consisting of a series of Gabor-elements whose properties were either derived from images of natural objects and fractals or manipulated to create simple parallel or circular patterns. All four stimulus classes used exactly the same number of Gabor elements distributed identically across three spatial-frequency bands, and differed only in the higher level structure emerging from the constituent elements' position (Figure 1B). The classes were selected such that they varied in the complexity of their orientation structure, with complexity being inversely related to the predictability of the orientation of one Gabor element given the orientation of another Gabor element in proximity. Parallel and circular patterns possess less complex (more predictable) orientation structure compared to object and fractal patterns. In addition, fractal patterns were unfamiliar to observers compared to the object stimuli, while retaining comparably complex orientation structure. We hypothesized that the difference in familiarity and predictability among stimulus classes would lead to different orientation sensitivity thresholds for the different image classes.

A pedestal-plus-test paradigm was used to assess the just-noticeable difference (JND) in orientation SD for different levels of added external noise within each stimulus class (Figure 1C). We found that the JND improved with the increase of pedestal noise up to an intermediate level, producing a characteristic dipper function (Solomon, 2009), and that the shape of this function varied with the stimulus class being viewed,
even though at low-level characteristics, these classes were fully matched. We argue that traditional models of vision that assume a feed-forward transfer of information from early sensory regions to areas involved in later stages of cognition are unable to account for the stimulus-class dependence we observed in our measurements. We propose an alternative probabilistic model, in which coding of orientation information is dynamically context dependent even at early stages of visual perception, and this context is modulated not only by low- (Schwartz, Hsu, & Dayan, 2007) but also high-level information.

Materials and methods

Participants

Eight undergraduate students from Central European University with normal or corrected-to-normal vision participated for monetary payment. All participants were unaware of the goals of the study.

Stimuli and apparatus

Our stimulus generation technique was inspired by Kingdom, Hayes, and Field (2001), who generated noise patterns composed of large numbers of randomly positioned Gabor elements. With these stimuli, it is possible to control the distribution of power and the amplitude spectra by varying either the number or the contrast of the individual Gabor elements. We extend this idea further by controlling the contrast, position, phase, and orientation of each element. This process allows us to generate synthetic natural images as stimuli over which we have more control than is possible with unprocessed natural images. Forty object and fractal images were selected from a database, scaled to $256 \times 256$ pixels, and then analyzed with a bank of steered wavelet filters (Freeman & Adelson, 1991) defined as

$$F_\theta = \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y} \exp \left( -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} + i(\omega_x s_0) \right).$$

Spatial frequency $\omega_x$ was 4, 8, and 16 c/° (wavelength: $\lambda_x = 32, 16, 8$ pixels, respectively), orientation $\theta$ was spaced at 45° intervals from 0° to 135°, $s_0 = x\cos\theta + y\sin\theta$, and the standard deviations of the Gaussian envelope were $\sigma_x = \sigma_y = 0.25\lambda_x$. The images were analyzed separately at each spatial frequency, and the response magnitude $R$ of each filter at each point in the image is

$$R_{(x,y)} = \sum_{\theta=0}^{135} \sqrt{\text{real}_{(x,y)}^2 + \text{imag}_{(x,y)}^2}.$$  \hspace{1cm} (2)

The phase $\phi$ at each point in the image is

$$\phi_{(x,y)} = \text{atan2} \left( \text{real}_{(x,y)}, \text{imag}_{(x,y)} \right).$$  \hspace{1cm} (3)

The interpolated orientation $\hat{\theta}$ at each point in the image is calculated as

$$\hat{\theta}_{(x,y)} = \text{atan2} \left( \sum_{\theta=0}^{135} R_{(x,y)} \sin(\theta), \sum_{\theta=0}^{135} R_{(x,y)} \cos(\theta) \right).$$  \hspace{1cm} (4)

Prior to the experiment, this analysis was performed on each image, yielding a pixel-by-pixel record of the local amplitude, phase, and orientation information of the image for each spatial frequency. Next, this pixel-by-pixel record was used for each image individually to generate a prespecified number of Gabor elements, starting from the highest amplitude pixel-by-pixel record and using the same Michelson contrast for each element. Michelson contrast is defined as

$$C_M = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}},$$

where $L_{\text{max}}$ and $L_{\text{min}}$ refer to maximum and minimum luminance, respectively. Finally, summing these Gabor elements composed our synthetic natural image.

Each Gabor element within the image was selected with a constraint of minimum center-to-center spacing of 2ε between adjacent elements to avoid overlaps and to match the spacing in the four stimulus classes (Figure 1B). For parallel patterns, the orientation of each element was drawn from a normal distribution with the same mean orientation within one trial and selected randomly from a normal distribution across trials. For circular patterns, the orientation of each element was drawn from a normal distribution with a mean orientation determined by its position relative to the center of the image, $\theta_{(x,y)} = \text{atan2}(-y, x)$, to produce a circular global pattern. For all stimulus classes, the total number of elements at each spatial frequency was fixed at 400 elements at 8 c/°, 200 elements at 4 c/°, and 100 elements at 2 c/°, ensuring that on average, all images had the same RMS contrast and that the RMS contrast across octaves was the same. Furthermore, all images had a 1/f amplitude spectrum that is close to that found in natural images (Field, 1987).

Stimuli were generated in MATLAB using custom software that incorporated routines from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), and presented on a Dell 1907FP 19-in. monitor with a mean luminance of 50 cd/m² and a refresh rate of 60 Hz. The display measured 36° horizontally (1152 pixels) and 27°
vertically (870 pixels), and was 57 cm from the observer, in a semilit room. The RGB monitor settings were adjusted so that the luminance of green was twice that of red, which in turn was twice that of blue. This shifted the white point of the monitor to 0.31, 0.28 (x, y) at 50 cd/m². A bit-stealing algorithm (Tyler, 1997) was used to obtain 10.8 bits (1,785 levels) of luminance resolution on the RGB monitor.

**Procedure**

On each trial, observers first viewed a noiseless reference image for 800 ms above a fixation point (Figure 1C). After a 500-ms delay with a gray background screen, a pedestal and a staircase image appeared together for 800 ms below the fixation point, randomly alternating sides across trials. The pedestal image contained all the Gabor elements of the reference image, but the orientation of each was randomly perturbed by a pedestal orientation noise drawn from a normal distribution with zero mean and an SD of $\sigma_p$, where $\sigma_p$ was one of six possible values between 1 and 32 in log steps. The staircase image was identical to the pedestal image, except that the SD of the normal distribution from which the orientation of each element was drawn was increased by $\Delta \sigma$ under the control of a staircase (Barlow, 1956). The pedestal noise manipulations altered the orientation of 99% of the elements by less than ±90° so that wrapping of orientations did not become an issue, even though a slightly higher percentage wrapping could occur due to the addition of staircase noise (average wrapping at the highest pedestal level was 5.75%).

The observer’s task was to make a two-alternative forced-choice decision whether the noisier (staircase) image was on the left or the right. Visual feedback was provided with a photometrically isoluminant fixation mark, which was green following a correct response and red following an incorrect response. All stimuli were presented with raised cosine spatiotemporal envelopes. The spatial windows were circular and subtended 4° in radius with edges smoothed over 0.5°. Image contrast was ramped on and off over three video frames (50 ms). All four stimuli classes were used in each experiment, blocked and counterbalanced across observers. Within each block, the pedestal noise levels were randomly interleaved trial by trial. Prior to the actual experiment, observers were familiarized with the stimulus material and experimental procedure by completing 70 practice trials.

**Data analysis and modeling**

JNDS for orientation SD were estimated from the 75% correct point of a cumulative normal psychometric function that was fitted to data from each experimental condition by a maximum-likelihood routine (Fründ, Haenel, & Wichmann, 2011). To produce 95% confidence intervals, 2,000 data sets were created using the fitted parameters as generating parameters and assuming a binomial distribution of trials.

We fitted a two-parameter model to the data incorporating a modified Weber function along with a “hard” threshold $\Theta$ (Bex, Mareschal, & Dakin, 2007) that represents an internal level of orientation noise discount (see Results):

$$\Delta \sigma_0 = \begin{cases} \sqrt{w \Theta^2 + \Theta^2 - \sigma_p^2} & \text{for } \Theta^2 > \sigma_p^2 \\ \sqrt{w \sigma_p^2} & \text{for } \Theta^2 \leq \sigma_p^2 \end{cases}$$

Here $\sigma_p^2$ is the pedestal orientation noise in orientation variance, $\Delta \sigma_0$ is the JND in orientation SD, and $w$ is the Weber fraction for orientation variance at pedestal levels exceeding the threshold. The parameters of the model were determined with least-squares minimization, weighted with SDs obtained by bootstrapping, from the estimated JND at each pedestal level and stimulus type.

The fit was compared with that of a two-parameter noisy inefficient (but otherwise ideal) observer model for orientation variance discrimination, introduced by Morgan, Chubb, and Solomon (2008), in which a fixed level of intrinsic noise determines the threshold for variance computation in an image:

$$P(c) = 1 - F \left[ \frac{\sigma_p^2 + \sigma_{int}^2}{(\sigma_p + \Delta \sigma)^2 + \sigma_{int}^2}, d_1, d_2 \right].$$

The model relates increment in staircase orientation noise $\Delta \sigma$ to expected response accuracy $P(c)$, given a pedestal level of orientation noise $\sigma_p$, an intrinsic noise level $\sigma_{int}$ as a free parameter, and two degrees of freedom $d_1$ and $d_2$ for the F-distribution. The degrees of freedom reflect how many individual elements an observer used for estimating the variance, and they are both equal to the second free parameter $M$ ($d_1 = d_2 = M$) of the model. The parameters of the model were estimated with a maximum-likelihood fit to data vectors consisting of the pedestal noise level, added staircase noise, stimulus class, and the observer’s response on each trial.

The coefficient of determination $R^2$ was used to estimate and compare the overall goodness of fit of the two models. Identifiability of the models was assessed from the mutual parameter covariance matrices of each model (Walter & Pronzato, 1996).

We developed a probabilistic model, inspired by the hard-threshold model, to provide a quantitative and descriptive account of how the observed human behavioral performance can be captured. A key aspect of the model is that it evaluates the visual input in the
context of the model’s internal representation of the task, which is implemented through prior distributions of the different image classes as learned through life experience. We phrased our probabilistic model in terms of inferring the average squared difference (orientation noise) $\sigma^2$ between orientations of segments in an image and a template assumed to be available a priori. Probabilistic inference yields a posterior $P(\sigma^2_0 | D)$ that describes how probable a given variance estimate is when observing the data $D$, where $D$ is defined as $D = \{\Delta x_1, \ldots, \Delta x_n\}$, the set of differences in orientation $\Delta x_i$ between the observed and expected orientations of each segment in the image, and where the expected orientations are defined by the a priori template. The inferred posterior could be used to compare the pedestal and staircase images when making a decision about whether one or the other image has a higher noise level. Segments of the image were assumed to be independently and identically distributed; therefore the likelihood of the data is the product of the individual orientation deviations from the template:

$$P(D | \sigma^2_0) = \prod_i P(\Delta x_i | \sigma^2_0).$$

(9)

We assume that the probability of the orientation difference $\Delta x_i$ of a segment and the template can be described by a zero-mean normal distribution:

$$P(\Delta x_i | \sigma^2_0) = \text{Normal}(\Delta x_i; 0, \sigma^2_0).$$

(10)

In this formulation, we omit the observer’s measurement noise for the sake of simplicity. While measurement noise influences the final discounting of orientation variance, it is nonspecific with regard to stimulus class. As a consequence, any measurement noise in this model is implicitly embedded in the a priori template.

In order to calculate the posterior distribution $P(\sigma^2_0 | D)$, we assumed that the prior of $\sigma^2_0$ is the conjugate prior of the normal distribution, which is an inverse gamma distribution. The choice of a conjugate prior helps us to derive the posterior in a closed form of another inverse gamma distribution. The mean of the prior indicates how strict or loose the model’s prediction of orientation variance is, on average, for one particular stimulus class (see Discussion for why stimulus classes differ in this respect). We characterized the prior with its prior hyperparameter for shape $\alpha'$ and prior hyperparameter $\beta'$ that scales the variance:

$$P(\sigma^2_0) = \text{InvGamma}(\alpha', \beta').$$

$$= \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} (\sigma^2_0)^{-(\alpha' + 1)} e^{-\left(\frac{\beta'}{\sigma^2_0}\right)}.$$  

(11)

We further note that the mean of the prior is equal to the model’s prior assumed orientation variance $\sigma^2_{\text{model}}$, causing the prior hyperparameter for shape to be a function of that assumed orientation variance:

$$\sigma^2_{\text{model}} = \frac{\beta'}{\alpha' - 1} \iff \alpha' = \frac{\beta'}{\sigma^2_{\text{model}}} + 1.$$  

(12)

Using this factorization we can derive the posterior distribution, which becomes an inverse gamma function with the following updated posterior hyperparameters as arguments:

$$P(\sigma^2_0 | D) = \text{InvGamma}(\alpha = \alpha' + \frac{2}{M}, \beta = \beta' + \frac{\beta}{\sigma^2_{\text{model}}} + \frac{2}{M}).$$  

(13)

where $M$ is the sample size—and in this case the number of orientation segments present in the image. From Equation 13, the maximum a posteriori estimate is derived as the mode of the distribution:

$$\sigma^2_{\text{MAP}} = \frac{\beta}{\alpha + 1} = \frac{\beta' + \sigma^2_{\text{pedestal}}}{\beta'/\sigma^2_{\text{model}} + M/2 + 1}. $$  

(14)

As a final step, we implemented a model on which observers can base their decision in each trial. This decision model compares the difference between the estimated overall noise of the pedestal and staircase images and determines when the ratio of the maximum a posteriori estimates of these images exceeds a ratio termed $W_{\text{Bayes}}$:

$$W_{\text{Bayes}} = \frac{\sigma^2_{\text{MAP, staircase}}}{\sigma^2_{\text{MAP, reference}} \cdot \sigma^2_{\text{MAP, pedestal}}} = \frac{\sigma^2_{\text{MAP, (}\sigma_0 + \Delta \sigma')}}{\max[\sigma^2_{\text{MAP, (}\sigma_0 + \Delta \sigma')}, \sigma^2_{\text{MAP, (}\sigma_0)\sigma^2_{\text{MAP, (}\sigma_0)}]} = \frac{\sigma^2_{\text{MAP, (}\sigma_0 + \Delta \sigma')}}{\sigma^2_{\text{MAP, (}\sigma_0)\sigma^2_{\text{MAP, (}\sigma_0)}} = \frac{\sigma^2_{\text{MAP, (}\sigma_0 + \Delta \sigma')}}{\sigma^2_{\text{MAP, (}\sigma_0)\sigma^2_{\text{MAP, (}\sigma_0)}} = \frac{2\beta' + (\sigma_0 + \Delta \sigma)^2}{\sigma^2_{\text{model}} + M\sigma^2_{\text{model}}}.$$

(15)

where $\sigma = \max[\sigma^2_{\text{model}}, \sigma^2_{\text{pedestal}}]$. When the pedestal noise is lower than the prior assumed variance, the pedestal image is perceived the same as the reference image. In this case, $W_{\text{Bayes}}$ indicates how large the difference between the sum of the pedestal and staircase noise and the prior model noise needs to be for an observer to correctly identify the staircase image (see Figure 2A, B). On the other hand, when the pedestal noise is higher than the prior model noise, $W_{\text{Bayes}}$ is equivalent to a Weber fraction—i.e., how much extra staircase orientation noise is needed at each pedestal noise level to perceive a difference (see Figure 2C).

In summary, there were three parameters in the model. The hyperparameter $\beta'$ was set to a fixed value, and its alteration had no effect on the performance of the model. The parameters $W_{\text{Bayes}}$ and $\sigma^2_{\text{model}}$ were fitted with Equation 15 to the empirically observed thresholds $\Delta \sigma$ by a maximum-likelihood routine to each stimulus class. Next, we used this model with the actual
experimental setup, computing the model’s choices over a large number of simulated pedestal and staircase images. Finally, the theoretical threshold results were determined by identifying for each pedestal noise and stimulus class the $D_M$ that just exceeded $W_{Bayes}$ and therefore became noticeable.

Results

Empirical findings

With all stimulus classes, human sensitivity showed a characteristic dipper function: At the lowest level of pedestal noise ($\sigma_p = 1^\circ$), the JND in orientation SD was relatively high; it gradually fell with added pedestal noise; it reached a minimum at around $\sigma_p = 16^\circ$; and then it rose steeply with the further addition of pedestal noise ($\sigma_p = 16^\circ$: $M = 4.31$, $SE = 0.46$; $\sigma_p = 32^\circ$: $M = 9.54$, $SE = 1.01$; $\Delta M = 5.24$, $SE = 0.65$, $p < 0.01$; see Figure 3A brackets). This was true for all observers (Figure 3A) and stimulus classes (Figure 3B).

Comparing the effect of pedestal noise collapsed across stimulus classes confirmed a significant dipper shape in the data; thresholds at the lowest and highest pedestal noise values were significantly greater than those at median pedestal values (see brackets and corresponding statistical tests of the threshold levels across pedestal noise levels in Figure 3A).

We used a repeated-measures two-way ANOVA with stimulus class and pedestal noise level as within-observer factors to test the hypotheses that thresholds at different levels of pedestal noise would show significant differences and that this pattern would change significantly with stimulus class. We found a significant main effect of both pedestal noise, $F(5, 35) = 27.99$, $p < 0.001$, $\eta^2_p = 0.80$, and stimulus class, $F(3, 21) = 34.2$, $p < 0.001$, $\eta^2_p = 0.83$, as well as a significant interaction effect between the two factors, $F(15, 105) = 2.55$, $p < 0.01$, $\eta^2_p = 0.27$, which suggests a changing shape of the dipper across stimulus classes. A post hoc comparison with Bonferroni correction indicated that, across all pedestal noise levels, observers had significantly higher JNDS with objects, circular, and parallel stimuli compared to fractals (object vs. fractal: $D_M = -4.43$, $SE = 0.64$, $p = 0.001$; circular vs. fractal: $\Delta M = -5.37$, $SE = 0.64$, $p = < 0.001$; parallel vs. fractal: $\Delta M = -4.20$, $SE = 0.41$, $p = < 0.001$).

To establish baseline JND plateaus at low pedestal noise levels before the dip manifests, we analyzed the thresholds across only the first three pedestal noise levels. The thresholds decreased with added pedestal noise, reaching a minimum at around $\sigma_p = 16^\circ$ and then rising steeply with higher pedestal noise. This was true for all observers (Figure 3B) and stimulus classes (Figure 3C).

Figure 2. Cartoon of the relationship between pedestal and staircase noise relative to the stimulus-class-dependent threshold noise $\sigma_{model}$ and $W_{Bayes}$ of the Bayesian model. The brackets indicate the amount of staircase noise that achieves a $W_{Bayes}$ of 1.4, and below each bar plot is given the relationship for estimating $W_{Bayes}$. (A) When the pedestal noise is much lower than the prior model noise, a considerable amount of staircase noise is needed before an observer reaches the JND determined by $W_{Bayes}$—at small pedestals, this produces an approximate plateau. (B) When pedestal noise is increased but still below threshold, less staircase noise is needed to achieve the same $W_{Bayes}$—this produces the dip in the JND curve. (C) At even higher pedestal noise, when it is above the $\sigma_{model}$ threshold, the JND is identical to the amount of staircase noise that exceeds the pedestal noise by a Weber fraction.
levels ($\sigma_p = 1^\circ-4^\circ$) and tested with a two-way repeated-measures ANOVA, again using stimulus class and pedestal noise level as within-observer factors. We found a main effect of pedestal noise, $F(2, 14) = 28.94, p = 0.003, \eta^2_p = 0.56$, indicating that, indeed, the gradual decrease class with added pedestal noise persisted regardless of the actual stimulus class. In addition, there was a main effect of stimulus class, $F(3, 21) = 9.04, p < 0.001, \eta^2_p = 0.81$, and the Bonferroni-corrected post hoc pair-wise comparison showed a systematic progression in average JND magnitude with stimulus class (Table 1, second column), increasing from circular to parallel patterns then to objects and lastly to fractals. We found no interactions between the two factors of pedestal noise and stimulus class, $F(6, 42) = 0.82, p = 0.56, \eta^2_p = 0.11$.

Model fitting

To quantify how well different models captured human behavior, we compared the fits of the inefficient-observer and hard-threshold models by testing the significance of the correlation between the data and fits for each model. Figure 4 shows fits of both models for three representative observers (rows) and for each stimulus class (columns).

The coefficient of determination between the fitted and empirical JNDs was overall more significant for the hard-threshold model than for the inefficient-observer model (paired-samples $t$ test), $t(31) = 2.84, p < 0.01, d = 0.50$ (Figure 5A).

To further validate the use of each model for explaining the JNDs, we compared the covariance matrices of the two fitted parameters of each model

<table>
<thead>
<tr>
<th>Stimulus class</th>
<th>Mean JND (SE)</th>
<th>Pair-wise comparison</th>
<th>Mean difference (SE)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractal</td>
<td>9.62 (0.68)</td>
<td>Object</td>
<td>2.84 (0.70)</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parallel</td>
<td>4.02 (0.66)</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Circular</td>
<td>4.90 (0.62)</td>
<td>0.001</td>
</tr>
<tr>
<td>Object</td>
<td>6.78 (0.75)</td>
<td>Parallel</td>
<td>1.18 (0.49)</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Circular</td>
<td>2.06 (0.36)</td>
<td>0.004</td>
</tr>
<tr>
<td>Parallel</td>
<td>5.60 (0.70)</td>
<td>Circular</td>
<td>0.88 (0.46)</td>
<td>0.58</td>
</tr>
<tr>
<td>Circular</td>
<td>4.17 (0.42)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Baseline plateau of empirical threshold (just-noticeable difference [JND] in degrees) across all observers from the first three pedestal noise levels (pedestal $\sigma_p = 1^\circ-4^\circ$). Post hoc pair-wise comparisons after repeated-measures two-way ANOVA of JNDs show a systematic and significant progression of JNDs with stimulus class from the lowest JND, with circular, to the highest, with fractal stimuli.

Figure 3. Measured empirical thresholds for discrimination of orientation SD (JND) showing the characteristic dip for each observer (A) and stimulus class (B), respectively. The thick black line in both panels represents the mean of the aggregated thresholds across all stimulus classes and observers. (A) Each thin line represents an observer aggregated across all stimulus classes. The brackets with corresponding statistical test represent post hoc pair-wise comparisons of the aggregated thresholds (thick black line) between the different pedestal levels from a repeated-measures one-way ANOVA ($\Delta M$: difference in mean value). (B) Each colored line represents thresholds for each stimulus class aggregated across observers. The shaded region indicates the standard error.
(\(\sigma_{\text{int}}\) and \(M\) for the inefficient-observer model; \(\Theta\) and \(w\) for the hard-threshold model). Low parameter covariance indicates a better identifiability of the model parameterization (Walter & Pronzato, 1996). Overall, the hard-threshold model displayed lower parameter covariance than the inefficient-observer model (paired-samples \(t\) test), \(t(31) = 42.51, p < 0.001, d = 8.15\) (Figure 5B).

The hard-threshold model and its Bayesian account provided a larger \(R^2\) for all stimulus classes than the inefficient-observer model across observers. The difference was significant for the fractal (paired-samples \(t\) test), \(t(7) = 2.86, p = 0.02, d = 1.01\), and object (paired-samples \(t\) test), \(t(7) = 2.42, p < 0.05, d = 0.86\), stimulus classes, and trending significant for the parallel stimulus class (paired-samples \(t\) test), \(t(7) = 2.30, p = \ldots\)
0.055, $d = 0.81$. A less pronounced modulation in the underlying data of the parallel and circular stimuli classes due to a floor effect makes it hard to confirm true differences between the fitted models of these classes.

As the hard-threshold model provided a better overall fit of the human threshold data, this model was used to further characterize the dipper function. The size of the dip, corresponding to the level of improvement in sensitivity as a function of pedestal noise, was calculated from the fit by comparing the hard threshold $H$ with the minima of the fitted function. Confirming the pair-wise comparison results on the raw data (see Table 1), dip size measured in orientation SD was greater for stimuli of fractals ($r_{fractal} = 8.26$) and objects ($r_{object} = 7.24$) than for parallel ($r_{parallel} = 5.84$) and circular ($r_{circular} = 5.28$) patterns. These results corroborate a systematic, class-specific reduction in dip size from fractals to other stimulus classes.

The Bayesian account of the hard-threshold model (Figure 6B) confirmed these observations and provided a possible explanation for the paradox of class-specific changes to JNDS. The fitted parameter $\sigma_{model}$ showed an orderly progression from circular to fractal stimuli (see the second-to-last row of Table 2). This gradual increase confirmed that the constraining effect of the sensory input compared to the internal template diminishes as one moves to more abstract stimulus classes.

The location of the dip, determined by $\Theta$ for the hard-threshold model and $\sigma_{model}$ for the Bayesian account, systematically elevated toward higher values across classes, as confirmed with paired-samples $t$ tests: For parallel and circular stimuli, the two curves were nearly identical, showing no difference in dip position—$\Theta$: $t(7) = 1.84, p = 0.11, d = 0.65; \sigma_{model}: t(7) = 0.72, p = 0.49, d = 0.26$. The dip’s position was significantly higher for object images than for circular images—$\Theta$: $t(7) = 3.79, p < 0.01, d = 1.34; \sigma_{model}: t(7) = 4.02, p < 0.01, d = 1.42$—and trending significantly higher than for parallel images—$\Theta$: $t(7) = 2.03, p = 0.08, d = 0.72; \sigma_{model}: t(7) = 1.90, p = 0.09, d = 0.68$. In turn, for fractals it was still significantly higher than for object stimuli—$\Theta$: $t(7) = 3.11, p = 0.02, d = 1.10; \sigma_{model}: t(7) = 3.11, p = 0.02, d = 1.10$. This pattern confirms that the bottom-up sensory influence on the Gabor elements decreases due to the increasing effect of the internal template (see Table 2, Bayesian model).

Finally, the elevation of the plateaus at small pedestal noises, also determined by $\Theta$ of the hard-threshold model, showed the same orderly progress as the dip position: For parallel and circular patterns the plateaus were similarly low, but for objects it was significantly higher, and for fractals it was even higher than for objects (see Table 2), confirming the pattern of the empirical JNDS (see Table 1).

In summary, human performance in noticing small amounts of added orientation noise was found to be paradoxically better when the stimulus images were not almost perfect but rather already carried a relatively high level of orientation variability. This performance
was systematically modulated across different stimulus classes defined by complexity and familiarity. The hard-threshold model assuming experience-dependent thresholds and its probabilistic Bayesian account could capture these characteristics of human performance better than the inefficient-observer model that assumes variance estimation from purely bottom-up processing.

**Discussion**

Finding a dipper function in human sensitivity performance is not without precedent in visual psychophysics (Solomon, 2009). Similar dippers in threshold have been found previously in many pedestal-based studies testing sensitivity to either mean or variance of visual attributes. For example, dippers have been reported in mean contrast discrimination (Nachmias & Sansbury, 1974) and blur discrimination tasks (Watt & Morgan, 1983), as well as in studies measuring JNDs of orientation with regular texture patterns (Morgan et al., 2008; Motoyoshi & Nishida, 2001) and in measures of higher level phenomena such as facial identity (Dakin & Omigie, 2009). Despite large differences in the characteristics of the stimuli, our results are in good agreement with others reported in the literature. Specifically, even though qualitatively

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Description</th>
<th>Fractal (SE)</th>
<th>Object (SE)</th>
<th>Parallel (SE)</th>
<th>Circular (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inefficient observer</td>
<td>$\sigma_{\text{int}}$</td>
<td>Intrinsic noise</td>
<td>15.03 (1.87)$^O$</td>
<td>17.86 (2.25)$^{PC}$</td>
<td>9.88 (0.78)$^{FO}$</td>
<td>10.78 (1.12)$^{IO}$</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>Gabor variances</td>
<td>9 (2.64)$^O$</td>
<td>53 (17.34)$^P$</td>
<td>18 (4.73)$^P$</td>
<td>25 (5.12)$^P$</td>
</tr>
<tr>
<td>Hard threshold</td>
<td>$\Theta$</td>
<td>Hard threshold</td>
<td>9.62 (0.66)$^{OP,PC}$</td>
<td>7.34 (0.73)$^{FC}$</td>
<td>6.22 (0.57)$^F$</td>
<td>5.44 (0.29)$^{FO}$</td>
</tr>
<tr>
<td></td>
<td>$w$</td>
<td>Weber fraction</td>
<td>0.244 (0.061)$^{OP,PC}$</td>
<td>0.038 (0.008)$^{FC}$</td>
<td>0.102 (0.030)$^F$</td>
<td>0.054 (0.015)$^{FO}$</td>
</tr>
<tr>
<td>Bayesian</td>
<td>$\sigma_{\text{model}}$</td>
<td>Noise prior</td>
<td>9.64 (0.48)$^{OP,PC}$</td>
<td>7.17 (0.57)$^{PC}$</td>
<td>6.06 (0.51)$^F$</td>
<td>5.33 (0.40)$^{FO}$</td>
</tr>
<tr>
<td></td>
<td>$W_{\text{Bayes}}$</td>
<td>Bayes fraction</td>
<td>1.22 (0.04)$^{OP,PC}$</td>
<td>1.04 (0.01)$^P$</td>
<td>1.08 (0.02)$^F$</td>
<td>1.05 (0.01)$^F$</td>
</tr>
</tbody>
</table>

Table 2. Parameter estimates (with standard error of the mean given in parentheses) from the three compared models for each stimulus class. The inefficient-observer and hard-threshold models were both fitted to each individual observer. The Bayesian model was fitted across all observers to obtain grand-average parameter estimates. Both the hard-threshold model and the Bayesian account show a systematic progression in the estimated threshold parameters with stimulus class going from fractal to circular, in contrast to the estimated parameters of the inefficient-observer model. Notes: Superscripts indicate that the given parameter estimate is significantly different (paired-samples t test, $p < 0.05$) from the stimulus class indicated by the superscript. $F$ = fractal; $O$ = object; $P$ = parallel; $C$ = circular.
different, the closest to our parallel stimuli are the oriented, arranged texture patterns of Morgan et al. (2008). These authors reported a threshold function of orientation SD discrimination with a plateau around 3°–4° at low pedestal orientation noise and a dip around 4°–5° of pedestal noise where the orientation JND reduces to about 1°, which is in remarkable agreement with our result (cf. Figures 3 and 6).

Although they used facial stimuli and different tasks, Dakin and Omigie (2009) reported a dipper with significantly elevated values with more naturalistic stimuli. This is in line with our finding of an elevated sensitivity curve with enhanced and changing dipper characteristics as the stimulus gets more complex, even though complex face discrimination cannot be related directly to measures of sensitivity to the mean or the SD of low-level attributes such as orientation. The advantage of our approach is exactly its ability to treat selectivity to orientation noise in a unified framework across different stimulus classes with equated low-level characteristics. Any additional sources of complexity that change human visual performance consequently have to come from extra stored knowledge rather than from extra visual input from the stimulus.

In order to achieve our goal, we stepped away from traditional stimulus displays, presenting either individual or a small set of separated Gabor elements organized into a regular lattice (Morgan et al., 2008; Regan & Beverley, 1985) or following the outlines of natural objects (Sassi, Vancleef, Machilsen, Panis, & Wagemans, 2010). Instead, we composed stimuli of an equal number of sparsely positioned Gabor elements derived from classes of images (fractals, natural objects, and simple circular and parallel patterns), which allowed the different higher order structures and contextual information in those classes to be preserved. Our method ensured that all low-level characteristics were equated across the stimuli classes and that they only differed in high-level aspects, possibly recruiting different amounts of top-down processing.

As a result of our approach, we could investigate the specifics and the origin of the dipper and JND levels in a complex context more thoroughly than was possible before. In agreement with Morgan et al. (2008), we found that while humans are known to have access to fine orientation information when viewing a single grating (Regan & Beverley, 1985), in context their JNDS are higher. Moreover, this JND is lowest at intermediate levels of pedestal noise, producing a curve with a dipper. In addition, we also showed how specific classes of stimuli yield a systematic change in parameters determining the height of the baseline plateau JNDS, and the position and depth of the dip in the dipper curve.

The purpose of the present article was not to establish the exact mechanism of this dipper, which would require extensive empirical testing with pedestal noise levels encompassing it. Rather, it was to show that neither of the competing explanations of the origin of the dipper (discussed later) can handle the baseline-elevation phenomenon with stimulus classes that we found, and to offer an alternative view. Since the focus of our interest is coding orientation information in general, and it is unclear whether humans can distinguish between variance changes and mean changes in natural scenes (as opposed to artificial test stimuli), we consider proposed mechanisms for both types of dipper results when contrasting with our data. Traditionally, a number of explanations have been proposed for the origin of such dippers (Solomon, 2009). Two of the most prominent explanations are nonlinearity of the transducer function provided for mean discrimination (Legge & Foley, 1980; Morgan et al., 2008; Stromeyer & Klein, 1974) and the existence of a hard threshold in the system (e.g., Bex et al., 2007; Crozier, 1950). For nonlinear-transducer models, a dip occurs because sensitivity is highest where the slope of the transducer function is steepest (Solomon, 2009). Such transducer functions are evident for contrast, but no studies of neurons with an analogous transducer function representing orientation variance have been reported in the mammalian visual system, and thus this reasoning cannot explain our results.

In the alternative of hard-threshold models, early sensors do not encode any information outside their resolution, which is preset so that it discounts the imperfections caused solely by the system rather than carried by the input (Solomon, 2009). For example, in the case of blur, the early part of the human visual system does not code blur below a criterion level defined by the limits of the optics and spatial resolution of the eyes. As a result, humans are left unaware of any blur of the external stimulus unless it exceeds the criterion level of the system (Burr & Morgan, 1997). Similarly, Morgan et al. (2008) showed that even though without a hard threshold (which they call sensory threshold), the dipper of the data can be derived from the signal detection theory of variance discrimination of orientation, adding a hard threshold significantly improves the fit. However, intrinsic noise or an analogous low-level mechanism for setting an orientation hard threshold cannot explain our results of these JND curves changing with stimulus class, because all low-level information is equated for by our stimulus generation method. We posit that such stimulus-dependent changes of the hard threshold must be determined by factors more complex than a hardwired early neural mechanism proposed for explaining earlier dipper results. We also point out that due to the continuous texture patterns in our stimuli without well-separated individual Gabor elements, the $M$ parameter
reflecting sample size in the original formulation of the inefficient-observer model loses its direct interpretation.

Given these shortcomings of the earlier proposals, what framework could capture our results? We suggest that at any moment, the orientation threshold below which exact orientation information of a small image patch does not get encoded is determined by a combination of bottom-up and top-down factors, including the statistical properties of the image elements, midlevel correlations among representations, and familiarity and experience with the content of the scene. This is an extension of the traditional contextual account of orientation coding that focuses on more local and low-level influences (Schwartz et al., 2007). Our proposal assumes that these factors are integrated into a higher level perceptual template (the percept) that acts as a Bayesian prior during perception and defines the expected features at different points in the image. In this scheme, the orientation of one or more features (here, Gabor elements) constrains the expected orientation of features at other locations and thus influences the coding precision of those features. This influence depends on both the statistical structure of the image determined by bottom-up sensory factors and the viewer’s internal high-level representation of the stimulus, which is evoked during viewing and manifests itself in top-down factors. In the case of the parallel and circular patterns, the constraint due to the high-level representation is nonspecific: Knowledge of the orientation at one location only loosely influences the expected orientation at another location, and hence the bottom-up sensory stimuli can dominate. In the case of images of real objects, the top-down influences are relatively strong due to shapes and contours (Humphreys, Riddoch, & Price, 1997; Malach et al., 1995). Even though contours follow an orderly relation in natural scenes (Geisler, Perry, Super, & Gallogly, 2001), our experience of viewing varying exemplars of objects from different projections and under occlusion means that knowledge of the orientation at one location influences the expected orientation elsewhere in the image. As long as a roughly correct orientation of an element satisfies the global expectations of the object’s shape, the percept is judged veridical without accessing the exact orientation information suitable for precise analysis. With fractals, the constraint of bottom-up information is even weaker compared to objects, due to unfamiliarity with the global shape, even though the basic visual structure of the fractals and object images is similar—that is, requiring only midlevel contour-related constraints to be satisfied. This notion is supported by Neri (2014), showing that low-level orientation processing is affected by such a high-level phenomenon as semantics, which in our study can be related to familiarity.

Although this interpretation is somewhat speculative, it is in agreement with an earlier suggestion by Morgan et al. (2008), which is a direct generalization of the proposal by Schwartz et al. (2007) of implementing a smoothness Bayesian prior in order to enhance effectiveness of perceiving natural scenes, and it has an important functional consequence. Specifically, Morgan et al. (2008) pointed out that the effect of the hard threshold in their paradigm of looking at a texture stimulus with orientation noise is that the texture appears slightly more regular than it really is. They suggested that this effect could be viewed as imposing a Bayesian prior based on the knowledge that contours in natural scenes tend to be continuous. We extend this idea by suggesting that rather than a single smoothness constraint, the top-down prior integrates a wide range of effects and constraints, all originating from some expectation regarding various aspects of natural scenes, making the percept more regular—i.e., more easily interpretable. The functional consequence of this proposal is that rather than viewing perception as a fundamentally feed-forward process with occasional local hardwired constraints implemented along the way, it opens the way to fully probabilistic frameworks, where each perceptual variable is viewed as latent and inferable, and derived thresholds dynamically vary, making the perceptual process immensely more flexible.

As a remark, we point out that the dip found in the sensitivity curve to changes in orientations can be shown to disappear when units of variance are used instead of units of SD (Solomon, 2009). This is a simple consequence of the interaction between the squaring operation and the magnitude of the effect, and it transforms a smaller dipper in JNDs measured in SD into a steeper initial slope measured in variance. However, the main finding of the present study is not tied to the dip or to the ability of our Bayesian model to capture this dip (Figures 4 and 6). Rather, our main result is demonstrating the shifting characteristics of this curve across stimulus types and providing a feasible Bayesian model that can account for this finding. As a similar shift in the curves will still be present in the data when using variance instead of SD to measure the JNDs that show a dip, our findings point to a general characteristic of orientation coding. In this article, we kept the units of JNDs in degrees of SD merely out of convention, rendering our work directly comparable to earlier work by Morgan et al. (2008).

Returning to the original question of orientation coding in natural scenes, this framework challenges the idea that coding of orientation is a fixed feature-detection process, which constantly provides high-precision orientation information to subsequent visual routines.
In conclusion, the present results lend support to the idea that the appearance of real objects and scenes, even at the level of such basic attributes as local orientation, is the outcome of a complex interplay between the responses of low-level, feed-forward spatial sensory information and the expectations of higher level perceptual processes. The visual system interprets ongoing sensory inputs based on prior experience to generate a behaviorally relevant representation of the current environment. This processing effectively and involuntarily discounts not only sensory noise arising from the limited resolution of visual detectors but also noise or uncertainty that is expected from viewing familiar objects under natural conditions that include novel illuminations, projections, and occlusions. Together with recent physiological studies (Berkes, Orban, Lengyel, & Fiser, 2011; Gilbert & Sigman, 2007), the present results suggest a continuous tuning of the precision for all basic attributes in the primary visual cortex that constantly integrates top-down influences to represent only the relevant aspects of a scene.

Keywords: orientation encoding, perceptual template, a priori expectations

Acknowledgments

This work has been supported by the Danish Agency for Independent Research, Sapere Aude program for Independent Research (JHC), and NSF IOS-1120938, Marie-Curie CIG 618918 (JF).

Commercial relationships: none.
Corresponding author: József Fiser.
Email: fiserj@ceu.edu.
Address: Department of Cognitive Science, Central European University, Budapest, Hungary.

References


Tyler, C. W. (1997). Colour bit-stealing to enhance the
