Journal of Mathematical Psychology 56 (2012) 427–443

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

A method for analyzing the dimensions of preattentive visual sensitivity

Charles Chubb [a,*], Ian Scofield [a], Chuan-Chin Chiao [b], George Sperling [a]

[a] Department of Cognitive Sciences, UC Irvine, Irvine, CA 92697-5100, United States

[b] Department of Life Science, National Tsing Hua University, Taiwan

ARTICLE INFO

Article history:
Received 29 August 2012
Received in revised form
4 January 2013
Available online 10 February 2013

Keywords:
Texture perception
Preattentive vision
Histogram contrast analysis

ABSTRACT

A fundamental question in vision science is: Which physical differences in the visual input are spontaneously visible and which are not? At present this question has only been partially answered. We propose that spontaneously visible variations are coded in "field-capture channels" that compute statistics on the raw visual input and pass them on to higher level processes. We describe a psychophysical method for exhaustively deriving the sensitivities of perceptually-available field-capture channels and thereby determining the dimensionality of early visual processes. The description of the field-capture channels resident in human vision will take the form of a compendium of dimensions of preattentive visual sensitivity. Here we demonstrate a method for deriving this compendium. In particular, we apply the method in a domain of physical variation (textures defined by randomly scrambled mixtures of different gray levels) for which the experimental data are available. A simulation shows that the method can (1) determine the number of field-capture channels that are differentially sensitive to variations in the domain and (2) derive a set of basis functions of the space of physical variations to which those channels are sensitive.

1. Introduction

The aim of this paper is to provide a method for deriving the dimensionality of early visual processes and to illustrate the method with an example texture discrimination task. The method developed here is predicated on several assumptions about the early processes in visual spatial discriminations. We assume that the initial segmentation of the visual field antecedent to deriving object boundaries, locations, or identities uses a battery of fast, spatially parallel image transformations whose response images reflect "the amounts of various kinds of visual substances present in the image" (Adelson, & Bergen, 1991). It is convenient to imagine these substance-sensing transformations as being implemented in retinotopically organized neural arrays; in any one of these arrays all neurons apply the same computation to the visual input but at different locations in the retina. Thus each array works continuously like a movie camera to capture a "neural image" (Robson, 1980) mapping the changing distribution of a particular "visual substance". To reflect the rapid, spatially parallel nature of these neural image transformations, we call them *field-capture channels*. The high cost in neural resources of a single field-capture channel makes it likely that human vision has only a modest number, suggesting that we may be able to figure out how many there are and what they sense.

Field-capture channels serve many purposes. First, they are used to segment the visual field into regions seen as differing in quality. Because we experience these differences spontaneously without any effort of attention, the processes that produce them are often called "preattentive"; however, this term is misleading: field-capture channels often come under attentional control. For example, in deciding whether or not to trust a rock face with your boot as you hike, you need to extract a summary statistic from the image of the rock face that reflects the probability that the rock face will hold your boot without slipping. By contrast, a completely different summary statistic is appropriate for selecting the softest towel on a shelf. In each case, an effort of attention is required to combine information from different field-capture channels to extract from the given surface a visual summary statistic that reflects the suitability of the surface for the use in question. Indeed one very useful approach to studying field-capture channels is to ask: what decision statistics can people achieve when they are asked to judge which of two texture samples has the higher level of a specified target property (e.g., higher luminance mean or higher luminance variance)? General methods for experiments of this sort are described in Chubb (1999). For applications, see Chubb and Nam (2000), Chubb and Talevich (2002), Nam and Chubb (2000).

Field-capture channels also play a central role in search tasks. For example, we assume that a target in a search task can be efficiently detected only if the observer can combine information from his/her field-capture channels to achieve an "attention filter" that is selective for the target vs the distractors. We submit that for this to be possible there must exist at least one field-capture channel that is activated by the target more than by any of the distractors. From this perspective, field-capture channels can be viewed as identical to the "feature maps" hypothesized to mediate

* Corresponding author.
E-mail addresses: cfchubb@uci.edu, c.chubb@uci.edu (C. Chubb).

0022-2496/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2013.01.001

visual search (Treisman & Gelade, 1980). However, we prefer the term "field-capture channel" to "feature map" which suggests that the corresponding image transformation pinpoints the locations of distinct occurrences of some specific sort of "feature" (e.g., redness or verticality); on the contrary, the transformation achieved by a given field-capture channel may well yield graded responses to a wide range of continuously varying image properties that are likely to defy definition in terms of a specific, easily characterized feature. We also view field-capture channels as enabling selective attention to the spatial distribution of a given property. It is precisely this sort of attention that is hypothesized to mediate "guided search," (Wolfe, Cave, & Franzel, 1989). In our view, the "guiding attributes" (Wolfe & Horowitz, 2004) used to control guided search must be synthesized by combining information from the field-capture channels resident in human vision.

More generally, we are concerned here with visual processing that we hypothesize can be characterized by three consecutive processing stages: (1) field-capture channels, (2) a salience map, and (3) detection and decision processes that operate on the salience map. We can view the physical stimulus I as the function that assigns to each point (x, y) in the stimulus field at each time t the vector $I(x, y, t) = (S(x, y, t), M(x, y, t), L(x, y, t))$ comprising the S-, M- and L-cone activations produced by the light hitting point (x, y) at time t. A field-capture channel C takes the stimulus I as input and produces as output a function $C[I]$ that assigns a real-valued image statistic $C[I](x, y, t)$ to each point (x, y, t). We assume the image statistic $C[I](x, y, t)$ is computed (using a rule that is invariant across different spatial locations (x, y) and times t) from the values assigned by I to points in space near (x, y) across times shortly before t; this statistic reflects the amount of the specific "visual substance" sensed by C that exists in the stimulus I in the neighborhood of (x, y) at time t.

We assume further that the output of the field-capture channels in human vision provides the bottom–up input to all subsequent visual processes. Here, we are concerned specifically with the input to a general purpose "salience map". The salience map assigns to each point (x, y, t) a scalar that represents the importance of the point (x, y) at time t for current computational purposes. We conjecture that this salience map is used for a wide range of different purposes including search, spatial localization, configural judgments (boundaries, shapes), numerosity judgments, etc.

The purpose of the illustrative experiments described here is to determine (1) the number of field-capture channels that are differentially sensitive to a particular class of stimuli, and (2) the space of discriminations within this class of stimuli that these field-capture channels enable. In these experiments participants judge the orientation of a square wave that modulates between different types of visual texture. We assume the participant uses his/her salience map to make this judgment. Specifically, we assume the participant selects several field-capture channels that are differentially sensitive to the sorts of texture to be discriminated; then, the participant combines in his/her salience map the responses of these field-capture channels to produce a pattern in which the alternating bars of the square wave differ as strongly in salience as the participant can achieve. The participant then bases his/her orientation judgment on the pattern of activation in the salience map. In this conceptualization, the salience map is a structure for coordinating input from different field-capture channels; the salience map itself contains no information about the qualities of the textures to be differentiated. In this model, then, there is an important distinction between seeing a difference between a region and its background versus knowing the content that fills the region—content is not accessible to salience.

For the moment, we leave open the question of exactly *how* salience is computed from the field-capture channel outputs. Salience is typically assumed to be a non-linear function of the individual channel responses; for example, the salience of the difference between two textures is sometimes assumed to be the Minkowski length (for some exponent β) of the difference between the vector of channel responses produced by one texture versus the other. However, there are good reasons to mistrust the Minkowski length as a model of salience. First, the salience of a texture difference is likely to depend strongly on the attention state of the participant. For example, if a participant is searching for a patch of texture that is brighter than the background, it is likely that salience will be more influenced by differences in mean texture brightness than if he/she is searching for a patch of texture that differs from its background in, say, RMS energy. Second, although the Minkowski length seems like a natural way to compute salience, it predicts that for any two textures A and B, a patch of A on a background of B should be equally salient as a patch of B on a background of A. This condition often turns out to be false. We will make much weaker, more general assumptions about how salience is computed from the field-capture channel responses. Indeed, one of the main contributions of this paper is to show how to measure functions that reflect the sensitivity of field-capture channels without making strong assumptions about how salience is computed.

Here, we demonstrate the methods of investigating the properties of field-capture channels using texture discrimination tasks as an example. To force the observer to rely exclusively on his/her field-capture channels in performing such texture discrimination tasks (versus even more complex cognitive processes), we require observers to discern differences between textures across space in displays of duration $\frac{1}{4}$ s or less-displays that are too brief to permit eye movements within the presentation. In the context of a texture discrimination task, we assume that a target region is preattentively discriminated from a background only if the two regions differentially activate one or more field-capture channels. For example, if they differentially activate one or more cone classes, target and background will be discriminable due to differences in brightness or color.

Since the work of Julesz (1962, 1975, 1981) and Beck (1966, 1982), it has been recognized that human vision has field-capture channels sensitive to purely textural properties (see also Beck, Sutter, & Ivry, 1987; Graham, 1989; Gurnsey & Browse, 1989; Julesz & Bergen, 1983). For example, in Fig. 1, the texture elements (or texels) of the test patch on the left differ in orientation from those of the background but not in spatial frequency, whereas the texels of the patch on the right differ in spatial frequency but not in orientation; in each case the test patch pops out from the background. Whenever two textures spontaneously segregate, we interpret this to mean the textures differentially activate one or more field-capture channels.

Many models of preattentive texture segregation have been offered (e.g., Beck, Prazdny, & Rosenfeld, 1983; Bergen & Landy, 1991; Bovik, Clark, & Geisler, 1990; Caelli, 1985; Fogel & Sagi, 1989; Graham, 1989, 1991; Graham, Beck, & Sutter, 1992; Grossberg & Mingolla, 1985; Knutsson & Granlund, 1983; Landy & Bergen, 1991; Malik & Perona, 1990; Wilson, 1993). All propose that human vision embodies a number of field-capture channels sensitive to local pattern orientation and spatial frequency. Often the proposed field-capture channels use spatially local linear filtering followed by rectification. Julesz (1962) famously conjectured that all preattentive texture discrimination used linear filtering followed by a squaring nonlinearity. And indeed, spectral energy accounts well for many cases of preattentive texture discrimination (Bergen & Adelson, 1988); however, many counterexamples exist (Diaconis & Freedman, 1981; Julesz, Gilbert, Sheppand, & Frisch, 1973; Julesz, Gilbert, & Victor, 1978; Pollack, 1971a,b, 1972, 1973). Although such examples and the models proposed to explain them are suggestive, little progress has been made in actually discovering

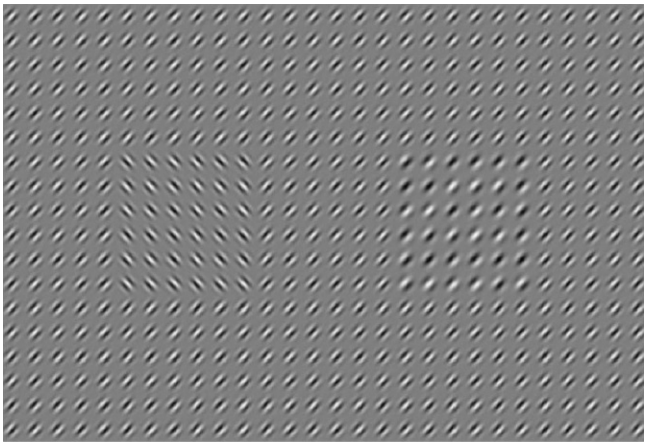

Fig. 1. Preattentive texture segregation driven by orientation (left) and spatial frequency (right).

and measuring the dimensions of preattentive visual sensitivity. The main obstacle has been lack of a systematic method. The point of this paper is to fill this void.

Recent innovations in texture synthesis algorithms offer a promising approach to these questions (Portilla & Simoncelli, 2000; Zhu, Liu, & Wu, 1999; Zhu, Wu, & Mumford, 1996). The general strategy taken in this work is to synthesize image samples equated in specified statistical properties to given target textures. The maximum entropy approach to this problem pioneered by Zhu et al. (1996), requires the synthesized image samples to be uniformly distributed within the space of all textures equated to the target texture in the specified statistics. This requirement is relaxed for the sake of computational efficiency by Portilla & Simoncelli; however, their algorithm plausibly approximates this ideal. In any case, for appropriately chosen statistical properties P, the synthesized image samples often appear remarkably similar to the target textures. (See Portilla & Simoncelli, 2000, for a fairly carefully pruned list of such properties P.) Any set P of properties that leads to synthesized samples that are always preattentively indiscriminable from their target textures must be seen as sufficient to describe human preattentive sensitivity. The difficulty has been in trying to distill such sufficient sets P of image statistics into a minimal, i.e. necessary, core set of statistics.

The methods we describe below promise to be useful in this effort. As we shall show, for specified subspaces of textures, these methods do indeed enable one to derive a set of statistics both necessary and sufficient to account for discrimination.

2. Groundwork

The methods we describe use psychophysical tasks that require the observer to discriminate one sort of texture from another. Many different discrimination tasks would work: for example, the observer might be asked to (1) identify the location of a target patch of one kind of texture in a background of another, or (2) identify the orientation of a square wave grating whose bars alternated between two different kinds of texture, or (3) indicate which of several distinct patches of texture was different from the others. For our purposes, the particular discrimination task is not important. What is important is the types of textures we use and how we manipulate the differences between them. These are described next.

2.1. Scrambles, histograms and perturbations

We always start with a particular set Ω of N_Ω different micropatterns. A micropattern is a "mini-image" that can serve as a single component in a large patch of texture; N_Ω is the number of micropatterns in Ω. Fig. 2 shows several possible micropattern sets. It is useful to distinguish three general classes of micropattern sets: linear, cyclic and unordered sets. The empirical questions of interest and the methods of analysis are likely to differ between these three different types of micropattern set. Fig. 2a and b are examples of linear micropattern sets. Fig. 2c is an example of a cyclic micropattern set, and Fig. 2d is an example of an unordered set. Linear and cyclic micropattern sets both conform to an obvious ordering, whereas unordered sets do not. On the other hand, linear sets have a unique pair of opposite, extreme elements whereas cyclic sets do not.

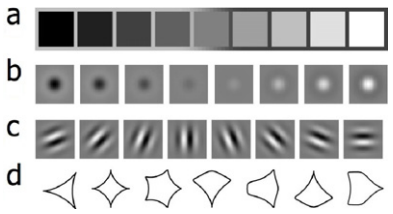

Fig. 2. Four different sets of micropatterns. (a) the linear set $\Omega_{Contrast}$, (b) another linear set, (c) a cyclic set, (d) an unordered set.

A scramble is a random texture composed of micropatterns whose frequencies in the scramble conform as exactly as possible to a particular probability distribution. To generate a scramble from a micropattern set Ω we first specify the proportions $p(\omega)$ with which different micropatterns $\omega \in \Omega$ appear in the scramble. Then we load a "virtual urn" with exactly the number of micropatterns needed to tile the stimulus region in exactly (or as nearly as possible) the proportions $p(\omega)$; then we draw from the urn randomly without replacement to assign micropatterns to the texel locations of the scramble. We call the resulting image an Ω-scramble and we call the probability distribution p the scramble histogram. Fig. 3 shows some examples of scrambles using micropattern set of Fig. 2a. The inset bar plots show the histograms.

We write $U(\omega)$ for the uniform histogram on Ω, i.e., the histogram that assigns probability $\frac{1}{N_\Omega}$ to all $\omega \in \Omega$ where N_Ω is the number of elements in Ω. The scramble shown in Fig. 3a has histogram U. By default we treat real-valued functions of Ω as a column vectors, writing $g^T h$ for the inner product of functions g, $h : \Omega \to \mathbb{R}$. That is

$$g^T h = \sum_{\omega \in \Omega} g(\omega) h(\omega). \tag{1}$$

We shall write $\|g\|$ for the Euclidean norm (or length) of g: $\|g\| = \sqrt{(g^T g)}$. Note that for any function $g : \Omega \to \mathbb{R}$, $g^T U$ is the mean value of g.

We call any real-valued function $\rho : \Omega \to \mathbb{R}$ a (histogram) *perturbation* if ρ sums to 0. Perturbations will play a central role in the method described below. Note that:

1. The difference between any two histograms is a perturbation.
2. Any perturbation ρ is orthogonal to U—i.e. ρ's mean value, $\rho^T U$, is 0.
3. $U + \rho$ is a histogram if and only if ρ is a perturbation whose minimum value is greater than or equal to $-\frac{1}{N_\Omega}$. $U + \rho$ and $U - \rho$ are both histograms if and only if the maximum absolute value of ρ is less than or equal to $\frac{1}{N_\Omega}$. For this reason, we call any perturbation ρ *maximal* if its maximum absolute value is $\frac{1}{N_\Omega}$. In addition, when we talk about the "maximal form" of a given perturbation ρ, we mean the perturbation $A\rho$ scaled to be maximal.

A central question addressed by the methods we develop is the following: For a given micropattern set Ω, what is the space of perturbations ρ for which scrambles with histograms $U + \rho$ vs. $U - \rho$ are preattentively discriminable?

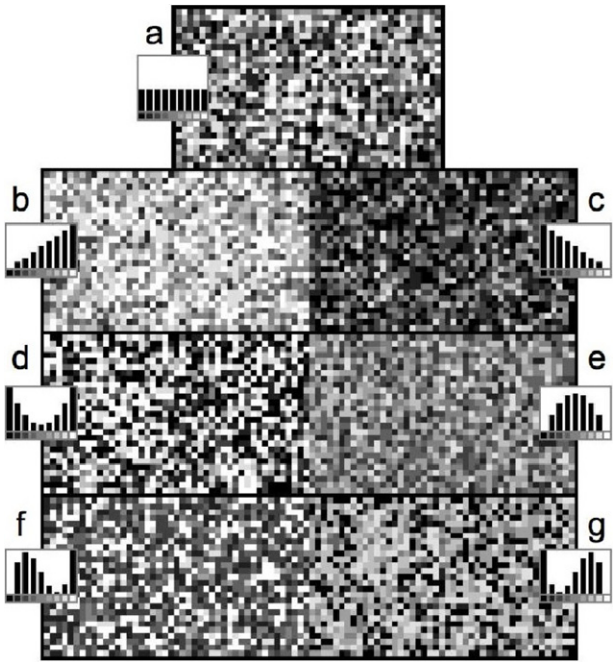

Fig. 3. Scrambles of $\Omega_{Contrast}$, with histograms (a) U (the uniform distribution; this assigns equal probability $\frac{1}{9}$ to all gray levels in Ω), (b) and (c) U plus complementary linear perturbations, (d) and (e) U plus complementary quadratic perturbations, (f) and (g) U plus complementary cubic perturbations, (h) and (i) U plus complementary 4th order perturbations. The linear, quadratic and cubic and 4th order perturbations used here are mutually orthogonal.

2.2. An example task

The methods we describe will work with many different discrimination tasks. We describe a simple task here to aid intuition. In this four-option, forced choice task, on every trial the observer is asked to try to detect the orientation of a square wave grating whose bars are filled with two different types of Ω-scramble. Specifically, for a number of reasons (discussed in Chubb, Econopouly, and Landy (1994)), it is useful to use scrambles that are perturbed away from the uniform histogram U in opposite directions. In other words, on every trial, we choose some perturbation ρ and set the histograms of scrambles to be discriminated equal to $U+\rho$ and $U-\rho$. Four example stimuli are shown in Fig. 4, one in each of the four possible orientations. The square waves in each of the two upper stimuli alternate between the two types of scramble shown in Fig. 3b and c. The lower left stimulus alternates between the two scrambles shown in Fig. 3d and e, and the lower right stimulus alternates between the two scrambles shown in Fig. 3f and g. On any given trial in our example discrimination task, the observer is given a brief display of a stimulus like one of the four in Fig. 4 and presses a button to indicate which of the four scramble-defined square wave orientations was present.

It will be convenient to use the terminology "ρ is discriminable with probability p" to indicate that the observer responds correctly with probability p when presented with a stimulus requiring the observer to discriminate scrambles with histograms $U+\rho$ vs. $U-\rho$. More generally, any statement about discriminating a perturbation ρ is to be taken as shorthand for the corresponding statement about discriminating scrambles with histograms $U+\rho$ vs. $U-\rho$.

2.3. A rough sketch of what is ahead

Our goal then is to determine the space of perturbations that human observers can discriminate. The general approach we shall take to this problem can be roughly sketched as follows. For a given "seed" perturbation ϕ, we engage the participant in a scramble discrimination task, where on each trial the perturbation ρ to be discriminated is highly correlated with ϕ. By thus fixing the predominant quality of the scramble difference the participant is asked to detect, we accomplish two related aims:

1. We enable the participant to optimize his/her performance for the perturbation ϕ.
2. We insure that across the restricted set of perturbations ρ tested, the visual salience of the difference between scrambles can be well approximated by a linear function of ρ.

We can then analyze how performance depends on the (relatively slight) deviations of the perturbations ρ away from ϕ; this enables us to derive a function $f_\phi : \Omega \to \mathbb{R}$ (called the "expansion" of ϕ) that reflects the influence exerted on the participant's judgments by each of the different types of micropatterns $\omega \in \Omega$ in the special circumstance in which the participant is striving to discriminate perturbations dominated by ϕ. By reiterating this procedure with a number of different seed perturbations ϕ_k, $k = 1, 2, \ldots, n$, we derive a set of expansions f_{ϕ_k}. As we shall show, if the seed perturbations ϕ_k are chosen appropriately, then the expansions f_{ϕ_k} will make up a basis of the space of all perturbations that human observers can discriminate. The relationship between this space and the field-capture channels resident in human vision is explored in the next section.

3. Differential sensitivity functions and spaces

3.1. Thinking about a single field-capture channel

We picture any field-capture channel M as being realized in the brain by a retinotopically organized array of neurons; that is, the neurons in M are arranged in the brain in an array, and the location in the array of a neuron reflects the location in the retina of the neuron's receptive field. Each neuron in the M-array uses the same (possibly complicated, nonlinear) computation to derive its response from the pattern of light in its receptive field. This computation defines the "visual substance" that M senses.

We assume that the visual substance sensed by M is at least moderately "fine-grained"; that is, the neurons in the M-array have relatively small receptive fields, and whatever lateral interactions occur between them are restricted in the scope of their influence. If so, then any neuron in M will be driven mainly by input in a small region of the retina, and the resulting neural image will carry relatively high acuity information. Otherwise, the neural image captured by M is likely to be too blurred to be useful.[1] When an Ω scramble stimulates the M-array, any given micropattern in the scramble is certain to influence the responses of multiple M neurons. On the one hand, neurons close to each other in M are likely to have overlapping receptive fields. Thus, wherever a micropattern $\omega \in \Omega$ occurs, it cannot avoid impinging upon the receptive fields of multiple M neurons. In addition, each micropattern is spatially extended; thus its impact spreads to include neurons whose receptive fields do not overlap.

[1] The term "fine-grained" is vague by design. It is difficult to anticipate which statistics are likely to possess important adaptive utility for human vision. For many purposes, high spatial resolution is likely to be important, suggesting a strong adaptive bias for field-capture channel receptive fields to be small. However, other factors are likely to predominate in determining the particular field-capture channels evolved by human vision. Consider, for example, a visual statistic whose value is high for image regions filled with the texture of a particular kind of grass favored by edible grubs. The spatial complexity of the statistic signaling this grass might well require large receptive fields; nonetheless, the adaptive benefit of evolving a field-capture channel selective for this statistic would be high. For reasons of this sort, we anticipate substantial variation in the sizes of the receptive fields of the field-capture channels in human vision.

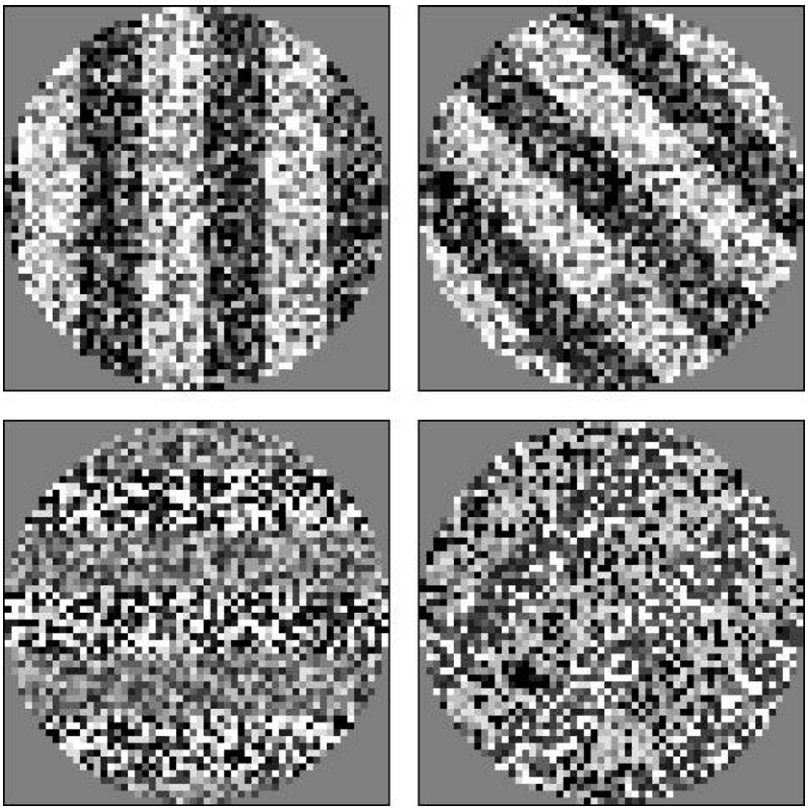

Fig. 4. Four example stimuli using the scrambles shown in Fig. 3. The task is a 4-option orientation judgment in which the possible orientations are illustrated in the four panels. The two upper stimuli alternate between scrambles of the types shown in Fig. 3b and c. The lower left stimulus alternates between the two scrambles shown in Fig. 3d and e, and the lower right stimulus alternates between the two scrambles shown in Fig. 3f and g.

Finally, M's neurons may laterally interact. If so, then a given micropattern can indirectly influence the activation levels of M-neurons by stimulating the receptive fields of M-neurons with lateral connections to them.

When an Ω scramble stimulates the M-array, different micropatterns of the same type are likely to produce different effects in M. The most important reason for this is that M's response to a given $\omega \in \Omega$ is likely to be somewhat context-dependent, and the contexts of two identical micropatterns within the scramble will almost always differ. For our purposes, all that matters about any such field-capture channel M are the average effects produced in M by different types of micropatterns $\omega \in \Omega$. We next formalize the concept of "first-order" sensitivity. For this purpose (and for much of what follows) we will need the following basic concept from linear algebra.

3.2. The projection of a function into a subspace

Let S be a subspace of the set of all real-valued functions of Ω. Let $b_1, b_2, \ldots, b_n$ be orthonormal functions spanning S, and let B be the matrix whose columns are the b_k's. Then for any function $g : \Omega \to \mathbb{R}$ define

$$g_S = \sum_{k=1}^{n} \left(b_k^T g\right) b_k = BB^T g \qquad \text{and} \qquad g_{\bar{S}} = g - g_S. \tag{2}$$

g_S is called the *projection* of g into S. By construction $g_S \in S$ and $g = g_{\bar{S}} + g_S$. In addition, it is easy to show that

1. for any function $h \in S$,
   (a) $g^T h = g_S^T h$, and
   (b) $g_{\bar{S}}^T h = 0$,
2. and conversely, for any function $\bar{h}$ orthogonal to S,
   (a) $g^T \bar{h} = g_{\bar{S}}^T \bar{h}$, and
   (b) $g_S^T \bar{h} = 0$.

For these reasons it makes sense to think of g_S as the portion of g that resides in subspace S and $g_{\bar{S}}$ as the residual portion of g orthogonal to S.

The concept of a projection is illustrated in Fig. 5. In this example, $\Omega = \{\omega_1, \omega_2, \omega_3\}$, and for any function $f : \Omega \to \mathbb{R}$, we use the vector notation "$f = (v_1, v_2, v_3)$" to indicate that $f(\omega_1) = v_1, f(\omega_2) = v_2$ and $f(\omega_3) = v_3$. In this example, $g = (2, 4, 5)$. We take S equal to the (1-dimensional) subspace of all $f = \alpha B$, for $\alpha \in \mathbb{R}$ and $B = (2, 4, 1)$. The subspace $\bar{S}$ is the plane (indicated by the inscribed rectangle) comprising all $f : \Omega \to \mathbb{R}$ orthogonal to B. The actual values of g_S and $g_{\bar{S}}$ are: $g_S = (2.38, 4.76, 1.19)$, and $g_{\bar{S}} = (-0.38, -0.76, 3.81)$. Note that $g_S \in S$, $g_{\bar{S}} \in \bar{S}$, and $g = g_S + g_{\bar{S}}$.

3.3. The differential sensitivity function of a field-capture channel

Returning to the field-capture channel M: for any element $\omega \in \Omega$, define $f_{total}(\omega)$ to be the mean activation produced in M by an occurrence of ω in a scramble. Let $\mu_{f_{total}}$ be the mean of $f_{total}(\omega)$ taken across all $\omega \in \Omega$. Define the *differential sensitivity function* of M with respect to ω as the function

$$f(\omega) = f_{total}(\omega) - \mu_{f_{total}}. \tag{3}$$

Note that the mean activation produced in M by a scramble with histogram p is

$$\begin{aligned} \sum_{\omega \in \Omega} f_{total}(\omega) p(\omega) &= \sum_{\omega \in \Omega} f(\omega) p(\omega) + \mu_{f_{total}} \sum_{\omega \in \Omega} p(\omega) \\ &= f^T p + \mu_{f_{total}}. \end{aligned} \tag{4}$$

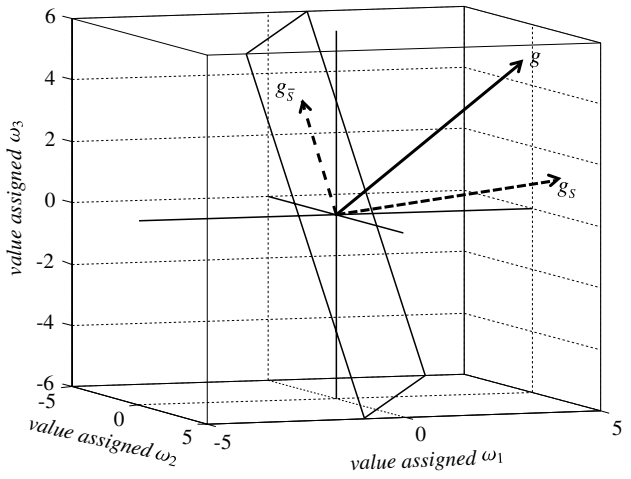

Fig. 5. Example illustrating the concept of a projection. In this example, $\Omega = \{\omega_1, \omega_2, \omega_3\}$. For any function $f : \Omega \to \mathbb{R}$, we write "$f = (v_1, v_2, v_3)$" as shorthand for "$f(\omega_1) = v_1, f(\omega_2) = v_2$ and $f(\omega_3) = v_3$". In this example, $g = (2, 4, 5)$; the (1-dimensional) subspace S is the set of all $f = \alpha B$, for $\alpha \in \mathbb{R}$ and $B = (2, 4, 1)$; $\bar{S}$ is the 2-dimensional subspace (indicated by the inscribed rectangle) comprising all $f : \Omega \to \mathbb{R}$ orthogonal to B; $g_S = (2.38, 4.76, 1.19)$, and $g_{\bar{S}} = (-0.38, -0.76, 3.81)$. Note that $g_S \in S, g_{\bar{S}} \in \bar{S}$, and $g = g_S + g_{\bar{S}}$.

Thus the difference in mean activation produced in M by scrambles with histograms p and q is $f^T(p - q)$, and in particular, for any perturbation ϕ, the difference in mean M activation produced by scrambles with histograms $U + \phi$ vs. $U - \phi$ is $2f^T\phi$.[2]

3.4. The differential sensitivity space and the null space of a micropattern set Ω

In general we expect human vision to have multiple field-capture channels that are differentially sensitive to the different ω's in a given micropattern set Ω. Let $f_1, f_2, \ldots, f_{N_{FCC}}$ be the differential sensitivity functions of these field-capture channels. (We shall be consistent in using the symbol N_{FCC} for the number of field-capture channels in human vision differentially sensitive to a given micropattern set Ω.) Then we call the f_k's the *differential sensitivity functions* of Ω, the space they span the *Ω differential sensitivity space* and the matrix F with column k equal to f_k the *differential sensitivity matrix* of Ω.

In addition, we call the space of perturbations orthogonal to all of the f_k's the *Ω null space*. Thus the union of the Ω differential sensitivity space and the Ω null space is the space of all perturbations.

For any perturbation ϕ, the kth entry of the vector $2F^T\phi$ gives the difference in mean activation produced in the kth field-capture channel by scrambles with histograms $U+\phi$ vs $U-\phi$. We call $2F^T\phi$ the *difference vector* produced by ϕ. Note that if ϕ is in the Ω null space, then $2F^T\phi = 0$.

3.4.1. The analogy to color perception

Let f_S, f_M and f_L be the functions characterizing the sensitivities of the S-, M- and L-cone classes to quanta of different wavelengths. Then for R, the matrix whose column vectors are f_S, f_M and f_L, the vector comprising the activations produced in the S, M and L cones by a light with spectrum P is given by $A_P = R^TP$. (Thus, the three components of A_P are the integrals (approximated as sums) across the continuum of visible wavelengths λ of $f_S(\lambda)\times P(\lambda), f_M(\lambda)\times P(\lambda)$ and $f_L(\lambda)\times P(\lambda)$.) Lights with spectra P and Q produce distinct color impressions only if $R^TP \neq R^TQ$. Conversely, lights with physically distinct spectra P and Q appear identical if $R^TP = R^TQ$, in which case the two lights are called *metamers*. Note that in this case, the function $D = Q - P$ is orthogonal to the space spanned by f_S, f_M and f_L because $R^TD = 0$.

We can view the discrimination of scrambles as analogous to the discrimination of homogeneous lights. Within this conceptual framework, the histogram of a scramble is analogous to the spectrum of a light, with one caveat: the histogram of a scramble is constrained to sum to 1 (because it is a probability distribution); however, the spectrum of a light need obey no such constraint. If we imagine a space of lights whose spectra are constrained to sum to a particular fixed value, then the analogy is exact: lights correspond to Ω-scrambles; their spectra correspond to scramble histograms; micropatterns in Ω correspond to quanta of different wavelengths. Just as lights with spectra P and Q are discriminable only if $R^TP \neq R^TQ$, scrambles with histograms p and q are discriminable only if $F^Tp \neq F^Tq$ for F the differential sensitivity matrix of Ω.

Analogously, if two Ω-scrambles are indiscriminable, we call them metamers and take this to indicate that the perturbation ϕ derived by taking the difference between their histograms has inner product zero with the sensitivity function of each field-capture channel that is differentially sensitive to the micropatterns in Ω; this means ϕ has a zero projection into the Ω-differential sensitivity space. And conversely, if two Ω-scrambles are discriminable, then we assume the difference ρ between their histograms has a nonzero projection into the Ω-differential sensitivity space. This basic fact provides the crucial leverage for the method described below.

3.4.2. Matching methods

Maxwell (1855) demonstrated the three-dimensionality of human vision using matching experiments. To mix the colors in his experiment, he used a flat-topped top pie-sliced into 100 sectors. Then when the top is spun, "the sectors of the different colours become indistinguishable, and the whole appears of one uniform tint". (Maxwell, 1855, p. 275.) To enable comparison of two different color mixtures, Maxwell used a smaller color-mixing disk inside a larger one; this produced an annulus of one tint hugging a central disk of another tint when the top was spun. In most of his experiments, Maxwell used varying proportions of three colors in the annulus (e.g., vermillion, emerald green and ultramarine) and two in the central disk, (e.g., black and white). He would have his participants look at the top only when it was spinning and would adjust the proportions of different colors in the central disk and in the surrounding annulus based on the feedback provided by his participants. In many cases, the participant was trying to achieve a spinning top that appeared uniformly gray. If, for example, the participant reported that the annulus appeared greenish and lighter than the center, Maxwell would (1) decrease the number of green sectors in the annulus, swapping them for blue and red sectors and (2) increase the number of white sectors in the central disk, swapping them for black sectors.

It is possible to use matching methods to investigate the dimensionality of sensitivity to gray-scale scramble variations. In fact, Chubb, Nam, Bindman, and Sperling (2007) used a matching method to show that the space of grayscale scrambles is perceptually three-dimensional. In these experiments, participants viewed square wave stimuli whose bars alternated between scrambles with histograms $U + \rho$ vs $U - \rho$, where on each adjustment trial ρ was a maximal perturbation. In different experimental conditions, Chubb et al. (2007) varied the basis B (analogous to the basis of colors used by Maxwell in a given mixture condition) used to generate ρ, and the participant sought to adjust the vector w of

[2] Note that $f_{total} - f$ (the portion of f_{total} not contained in the space of all perturbations) is a constant function—i.e., a function that assigns the same value to all $\omega \in \Omega$; only constant functions are orthogonal to all perturbations. Our methods do not allow us to estimate this constant component of f_{total}.

weights used to generate $\rho = Bw$ so that the square wave alternating between scrambles with histograms $U+\rho$ vs $U-\rho$ became as nearly invisible as possible. This minimally salient, maximal perturbation ρ was then tested in a forced-choice experiment in which the participant strove to judge the orientation of a square wave whose bars alternated between scrambles with histograms $U+\rho$ vs $U-\rho$. Chubb et al. (2007) tested four 3-dimensional subspaces (spanning different subspaces of 4th order polynomial perturbations) as well as the (4-dimensional) subspace of all fourth-order polynomial perturbations. For each of the 3-dimensional subspaces tested, the minimal salience perturbation yielded discrimination performance significantly greater than chance; however, the minimally salient, maximal perturbation in the subspace of all fourth-order polynomial perturbations yielded chance performance (within measurement error).

In the case of color perception, Maxwell had strong *a priori* reasons to think the space of homogeneous lights was perceptually three dimensional. As his text suggests, Maxwell presumed that any particular feedback from the observer in his experiment, e.g., "the annulus appears too green", could be canceled by an appropriate adjustment of the segment angles subtended by the base colors he used in his top. His experiment verified this intuition. Similarly, the experiments of Chubb et al. (1994) and Chubb, Landy, and Econopouly (2004) study had shown that the space of grayscale scrambles had perceptual dimension of at most three but had not ruled out the possibility that there might be only two dimensions of sensitivity. Chubb et al. (2007) were able to show that there existed three-dimensional subspaces of perturbations that contained no perturbations ρ orthogonal to the sensitivity functions of all the field-capture channels resident in human vision; this implies that human vision contains at least three field-capture channels sensitive to grayscale scrambles. In both of these cases, then, strong hypotheses existed prior to the research that might be tested directly using a matching paradigm.

In the absence of strong prior hypotheses, however, matching methods are likely to be of limited utility in analyzing the perceptual dimensionality of spaces of Ω-scrambles, especially in those cases in which $|\Omega|$ is fairly large. In any matching paradigm, we start by specifying some subspace S of perturbations. If the dimension of S is one greater than the dimension of the human sensitivity space for Ω, then there will exist in S at least one maximal perturbation ρ orthogonal to the human sensitivity space for Ω. In this case, the scrambles with histograms $U+\rho$ and $U-\rho$ will be preattentively metameric, and ρ should yield chance discrimination performance. We would like the matching paradigm to yield such a perturbation ρ if it exists in S.

Note that ρ is an element of the null space dual to the human sensitivity space for Ω. Of course, if we could determine the null space, then we would also know the human sensitivity space for Ω. However, when Ω is large, then the null space is likely to be of higher dimension than the sensitivity space, in which case it is likely to be easier to measure the sensitivity space directly using the methods described below.

4. Seed expansion

Seed expansion is a method for deriving a perturbation that is an element of the differential sensitivity space of a micropattern set Ω. By iterating this method several times, one can derive a minimal basis of the differential sensitivity space of Ω, thereby determining the dimensionality of this space.

4.1. Model assumptions

Any normalized perturbation ϕ whose maximal form yields near perfect discrimination can play the role of the "seed" in a seed-expansion task. The *correlation* between ϕ and any other perturbation ρ is the cosine of the angle formed between ϕ and ρ (when they are considered as N_Ω-dimensional vectors) which is given by

$$r(\phi,\rho)=\frac{\phi^T\rho}{\|\phi\|\|\rho\|}, \tag{5}$$

where $|\phi|$ and $|\rho|$ are the norms (the Euclidean lengths) of ϕ and ρ. In a seed-expansion task, on every trial the observer is asked to discriminate a perturbation ρ (variable across trials) that is highly correlated with ϕ, e.g., with correlation 0.9 or higher.[3]

Let F be the matrix whose column vectors are the (unknown) differential sensitivity functions f_k, $k=1,2,\ldots,N_{FCC}$ of the N_{FCC} field-capture channels enabling discrimination of Ω-scrambles. To perform the discrimination task given stimulus produced by a perturbation ρ, the participant must first combine the differential activations $f_k^T\rho$ in his/her salience map. We do not know exactly what computation the participant uses to generate this salience signal, and we would like the method we are developing to be robust with respect to potential variations in this computation. For this reason, we will impose very weak assumptions on the salience computation.

In the context of a task in which the observer is attempting to discriminate perturbations ρ all of which are highly correlated with a given seed perturbation ϕ, we assume that the salience of the difference between scrambles with histograms $U+\rho$ vs $U-\rho$ depends on the observer's attentional state which is determined by ϕ. For example, if ϕ were the perturbation used to generate the scrambles with histograms $U+\phi$ and $U-\phi$ shown in Fig. 3b and c, we would expect the participant to be selectively tuned in to differences in texture brightness. The participant might well use a different salience computation if ϕ were the perturbation used to generate the scrambles shown in Fig. 3d and e. Because we expect the salience computation to depend on the seed perturbation ϕ, we write "$Sal_\phi(\rho)$" explicitly subscripted by ϕ to indicate the salience of ρ in the context of the task in which the seed perturbation is ϕ.

How do the different field-capture channels combine to produce salience? The answer is, we do not know. For this reason we adopt a very general, weak model of salience that subsumes (we hope) all the plausible candidate computations as special cases. Specifically, we view the differential activations $f_k^T\rho$, $k=1,2,\ldots,N_{FCC}$ as the raw material to be combined. Then we admit the possibility that each of these pieces of raw material may by transformed by a real-valued function $G_{\phi,k}$. Note the implication carried by the subscript k that $G_{\phi,k}$ may be different for different field-capture channels, and by the subscript ϕ that $G_{\phi,k}$ may depend on the attention state adopted by the participant in performing the discrimination task when the seed perturbation is ϕ. (For example, we might expect that the participant selects the functions $G_{\phi,k}$ so as to emphasize the responses of those field-capture channels k that are most sensitive to variations in ϕ.) After using the functions $G_{\phi,k}$ to individually transform the outputs of the different field-capture channels, we add these transformed field-capture channel outputs together. Finally, we admit the possibility that this sum of transformed field-capture channel outputs may be further transformed by the application of

[3] The correlation level of 0.9 should be taken as a rough guideline. We have found that in practice, if the perturbations ρ all have correlation 0.9 or higher with the seed perturbation ϕ, then the model assumptions described below are typically satisfied within measurement error. On the other hand, if the correlations between the perturbations ρ and the seed perturbation ϕ are *too high*, then it becomes impossible to obtain a sharp estimate of the "expansion of ϕ" (defined below), the function reflecting the degree to which the participant's judgments are influenced by the different micropatterns $\omega \in \Omega$ when the participant is attempting to discriminate the seed perturbation ϕ.

a real-valued function H_ϕ that also may depend on that attention state of the participant as determined by the seed perturbation ϕ. In short, we assume that

$$Sal_\phi(\rho) = H_\phi\left(\sum_{k=1}^{N_{FCC}} G_{\phi,k}\left(f_k^T\rho\right)\right) \tag{6}$$

for differentiable functions $G_{\phi,1}, G_{\phi,2}, \ldots, G_{\phi,N_{FCC}}$ and H_ϕ. Note that Eq. (6) includes the Minkowski length as a special case. In particular, if $G_{\phi,k}(x) = |x|^\beta$ for all k, and $H_\phi(x) = x^{\frac{1}{\beta}}$, then $Sal_\phi(\rho)$ is the Minkowski length (with exponent β) of the vector $F^T\rho$ of differential field-capture activations.

Finally, we assume that probability correct is a psychometric function of $Sal_\phi(\rho)$. Any psychometric function that assigns chance probability to $x = 0$ would serve our purposes. For definiteness, we use a Weibull function: specifically, we assume that the probability that the observer correctly discriminates ρ is $\Psi(Sal_\phi(\rho))$, where

$$\Psi(x) = 0.25 + 0.75\left(1 - e^{-x^\gamma}\right). \tag{7}$$

There are two things to note about Eq. (7). First, this Weibull function assumes chance performance is 0.25 reflecting the fact that our example task is a 4-option, forced-choice task. Second, this Weibull function has only one free parameter, the exponent γ that controls the function's steepness. Typically, the Weibull function also includes a second parameter that enters as a divisor to the argument, x. This parameter will turn out to be redundant in our model because it can be absorbed into the parameters used to define $Sal_\phi(\rho)$.

Several important conceptual points are hidden behind the assumption that $\Psi(x)$ gives the probability correct on a trial in which the stimulus has salience x. On any given trial, the participant extracts from the stimulus a noise-corrupted statistic that he/she uses to make his/her decision. (If there were no noise in this process, then every time we presented a given stimulus, the participant would have to give the same response.) In our model, $Sal_\phi(\rho)$ is a deterministic (i.e., nonrandom) function of ϕ and ρ. Thus, all the randomness in the model enters via the psychometric function Ψ. The task we are using to illustrate the seed expansion method requires the participant to select one of four response options. It is natural to imagine that on a given trial, the participant computes four internal statistics, one for each option, and then selects the option for which the corresponding statistic is the largest. If one made specific assumptions about the distributional properties of these four hypothetical random variables (e.g., that they are independent, gamma distributed with parameters that depend in such and such ways on the stimulus), then one might be able to derive a formula for the cumulative distribution function $F_{decision}(x)$ characterizing the probability of a correct response given a stimulus with salience x. However, we do not need to solve this problem. We naturally expect $F_{decision}(x)$ to be a smooth, sigmoidally increasing function of x constrained to assign the value 0.25 (chance performance) to $x = 0$, and we are confident that we can approximate any such function tolerably well with a Weibull function of the form given in Eq. (7). Thus, the choice of the Weibull psychometric function is driven not by theoretical but by pragmatic considerations.

4.2. The gradient of salience

Let τ be the value of salience required to support a threshold probability correct p (i.e., $\tau = \Psi^{-1}(p)$ for Ψ given by Eq. (7)), and let α_τ be the positive scalar satisfying $Sal_\phi(\alpha_\tau\phi) = \tau$.

We shall assume that for all perturbations ρ sufficiently well-correlated with the seed perturbation ϕ, $Sal_\phi(\rho)$ can be well approximated by

$$\begin{aligned} Sal_\phi(\rho) &= Sal_\phi(\alpha_\tau\phi + (\rho - \alpha_\tau\phi)) \\ &\approx Sal_\phi(\alpha_\tau\phi) + \Delta_{\phi,\tau}^T(\rho - \alpha_\tau\phi) \\ &= \tau + \Delta_{\phi,\tau}^T(\rho - \alpha_\tau\phi), \end{aligned} \tag{8}$$

where $\Delta_{\phi,\tau} : \Omega \to \mathbb{R}$ is the gradient of Sal_ϕ at the point $\alpha_\tau\phi$. I.e., for any perturbation ρ,

$$\Delta_{\phi,\tau}^T\rho = \lim_{\epsilon\to 0}\left(\frac{Sal_\phi(\alpha_\tau\phi + \epsilon\rho) - Sal_\phi(\alpha_\tau\phi)}{\epsilon}\right). \tag{9}$$

The following theorem shows that if $Sal_\phi(\rho)$ conforms to Eq. (6), then $\Delta_{\phi,\tau}$ is an element of the differential sensitivity space of Ω and hence is eligible to serve as an element in a basis of the differential sensitivity space of Ω.

4.3. Theorem

Suppose $Sal_\phi(\rho)$ satisfies Eq. (6). Then $\Delta_{\phi,\tau}$ is a linear combination of the columns of F and hence an element of the Ω differential sensitivity space. Specifically, (using prime notation to denote the derivative of a function) for the scalar

$$c_{\phi,\tau} = H'_\phi\left(\sum_k G_{\phi,k}\left(\alpha_\tau f_k^T\phi\right)\right) \tag{10}$$

and the column vector $g_{\phi,\tau}$ of length N_{FCC} whose kth component is

$$g_{\phi,\tau,k} = G'_{\phi,k}\left(\alpha_\tau f_k^T\phi\right), \tag{11}$$

we shall show that

$$\Delta_{\phi,\tau} = c_{\phi,\tau}Fg_{\phi,\tau}. \tag{12}$$

Proof. It will be convenient to set

$$q_\rho(x) = Sal_\phi(\alpha_\tau\phi + x\rho). \tag{13}$$

Three applications of the chain rule to the right side of Eq. (13) yield

$$\begin{aligned} q'_\rho(x) &= H'_\phi\left(\sum_k G_{\phi,k}\left(f_k^T[\alpha_\tau\phi + x\rho]\right)\right) \\ &\quad\times \sum_k G'_{\phi,k}\left(f_k^T[\alpha_\tau\phi + x\rho]\right)f_k^T\rho. \end{aligned} \tag{14}$$

Note that

$$q'_\rho(x) = \lim_{\epsilon\to 0}\left(\frac{Sal(\alpha_\tau\phi + (x+\epsilon)\rho) - Sal(\alpha_\tau\phi + x\rho)}{\epsilon}\right). \tag{15}$$

Thus, in particular,

$$q'_\rho(0) = \lim_{\epsilon\to 0}\left(\frac{Sal(\alpha_\tau\phi + \epsilon\rho) - Sal(\alpha_\tau\phi)}{\epsilon}\right) = \Delta_{\phi,\tau}^T\rho. \tag{16}$$

On the other hand, it is easy to verify from Eq. (14) that $q'_\rho(0) = (c_{\phi,\tau}Fg_{\phi,\tau})^T\rho$. We conclude that for any perturbation ρ,

$$\Delta_{\phi,\tau}^T\rho = q'_\rho(0) = (c_{\phi,\tau}Fg_{\phi,\tau})^T\rho, \tag{17}$$

implying Eq. (12). □

4.4. The expansion of a seed perturbation

As above, for threshold probability p, suppose $\tau = \Psi^{-1}(p)$ is the salience supporting probability p correct, and let α_τ satisfy $Sal_\phi(\alpha_\tau\phi) = \tau$. For any of the perturbations ρ tested in the context

of an experiment to expand the seed ϕ, Eq. (8) leads to

$$P(correct|\rho) \approx \Psi(\tau + \Delta^T_{\phi,\tau}[\rho - \alpha_\tau \phi]). \tag{18}$$

It is a natural step from Eq. (18) to assume that $\Delta_\phi = \Delta_{\phi,\tau}$ is invariant with respect to τ. In this case, setting ρ to 0 implies $\tau = \alpha_\tau \Delta^T_\phi \phi$, leading to the simplification,

$$P(correct|\rho) \approx \Psi(\Delta^T_\phi \rho). \tag{19}$$

We call the function $\Delta_\phi : \Omega \to \mathbb{R}$ the *expansion* of the seed ϕ. The expansion Δ_ϕ reflects the influence exerted by different $\omega \in \Omega$ on the observer's judgments in the specific variant of the discrimination task in which the difference between the histograms of the scrambles presented on each trial is dominated by the seed perturbation ϕ. We call Δ_ϕ the "expansion" of ϕ to reflect the fact that although Δ_ϕ does indeed depend on ϕ and will undoubtedly be positively correlated with ϕ, it must also expand outward from ϕ to include contributions from all $N_\Omega - 1$ dimensions of the space of all perturbations.

Eqs. (19) and (7) define a model that can be used to derive a likelihood function. The parameters of the model are the exponent γ in (7) and the values assigned by function Δ_ϕ to the elements of Ω. The number of degrees of freedom is N_Ω because Δ_ϕ is constrained to sum to 0.

4.4.1. The likelihood function

Let ϕ be the seed whose expansion is to be derived. We assume that the perturbations ρ_t tested across all trials t are all well-correlated with ϕ. Then for any $\Delta_\phi : \Omega \to \mathbb{R}$ and any $\gamma \in \mathbb{R}$ the likelihood function for the model given by Eqs. (18) and (7) is

$$\Lambda_\phi(\Delta_\phi, \gamma) = \prod_{\text{Correct trials } t} \Psi(\Delta^T_\phi \rho_t) \times \prod_{\text{Incorrect trials } t} \left(1 - \Psi(\Delta^T_\phi \rho_t)\right) \tag{20}$$

where $\Psi(x)$ is defined by Eq. (7).

As we demonstrate in our simulation, it is easy to use the likelihood function given in Eq. (20) in the context of a Markov Chain Monte Carlo simulation to estimate Δ_ϕ and γ.

4.5. Iterating the seed-expansion method to derive a basis of the differential sensitivity space

As above, let F be the $N_\Omega \times N_{FCC}$ matrix whose column vectors are the differential sensitivity functions of Ω.

If human vision has any field-capture channels with substantial differential sensitivity to Ω, then there must exist some maximal perturbation that yields near perfect discrimination. We select such a perturbation and normalize it. We use this normalized perturbation ϕ_1 as our first seed and derive the expansion Δ_{ϕ_1}. Since Δ_{ϕ_1} is an element of the Ω differential sensitivity space, it can be taken as the first element of a basis of this space. We then test whether there exist any maximal perturbations orthogonal to Δ_{ϕ_1} that yield discrimination above chance. If not, we conclude that the differential sensitivity space of Ω is one-dimensional with basis $\{\Delta_{\phi_1}\}$. If such perturbations do exist, then we choose one. If this perturbation supports near perfect discrimination, then we normalize it to derive our second seed ϕ_2. If not, then we will need to construct a maximal perturbation comprising a weighted sum of the new perturbation with Δ_{ϕ_1} and adjust it to produce near perfect discrimination (e.g., a success rate of 0.9 or higher). We will then normalize this perturbation to derive our second seed, ϕ_2 and proceed to derive the expansion Δ_{ϕ_2}.

We reiterate this process as many times as needed: after step k we test whether the space of perturbations orthogonal to all the previously derived expansions, $\Delta_{\phi_1}, \Delta_{\phi_2}, \ldots, \Delta_{\phi_k}$, contains a maximal perturbation that yields discrimination significantly greater than chance. If not, we conclude that $\Delta_{\phi_1}, \Delta_{\phi_2}, \ldots, \Delta_{\phi_k}$ span the Ω sensitivity space. Otherwise we choose such a maximal perturbation ρ. If ρ supports near perfect discrimination, then we normalize it to derive the next seed ϕ_{k+1}. If not, then we construct a maximal perturbation $\hat{\phi}_{k+1}$ comprising a weighted sum of ρ and some perturbation known to produce near perfect discrimination, adjusting $\hat{\phi}_{k+1}$ to contain as much ρ as possible while producing near perfect discrimination (e.g., a success rate of 0.9). Then we will normalize $\hat{\phi}_{k+1}$ to obtain our next seed, ϕ_{k+1} and proceed to derive the next expansion $\Delta_{\phi_{k+1}}$.

This iterative procedure yields a basis $\Delta_{\phi_1}, \Delta_{\phi_2}, \ldots, \Delta_{\phi_n}$ of the Ω sensitivity space. It is possible, however, due to measurement error accrued step by step along the way, that this basis is not minimal. As a final step, we use singular value decomposition to extract the principal components of the matrix whose column vectors are the expansions $\Delta_{\phi_1}, \Delta_{\phi_2}, \ldots, \Delta_{\phi_n}$. We take as our final basis the set of those principal components whose eigenvalues are significantly greater than zero.

5. An example application

The purpose of this section is to illustrate how we actually derive a basis for a differential sensitivity space. As we proceed, we will address several important practical issues not discussed in the abstract description of the method above. In this section, we focus on textures composed of small squares of different Weber contrasts, as illustrated in Figs. 3 and 4.

5.1. The example field-capture channels

In order to illustrate the seed-expansion method, we need to pick a set of hypothetical field-capture channel sensitivity functions (which will become the columns of the matrix F); we also need to specify for any seed ϕ the functions H_ϕ and $G_{\phi,k}$, $k = 1, 2, \ldots, N_{FCC}$ used in Eq. (6) as well as the exponent γ in the psychometric function (Eq. (7)). Given these parameters, we will be able to simulate the seed-expansion method. In addition, we will be able to compare the estimates of the various expansions we derive with their actual values. To illustrate the iterative aspect of the method, this set of sensitivity functions should be of cardinality greater than one.

As our example set, we take three field-capture channels roughly consistent with previous research (Chubb et al., 1994, 2004, 2007). The sensitivity functions of these three hypothetical field-capture channels are shown in Fig. 6. For purposes of this paper, we will take these three functions to form a basis of the human differential sensitivity space for scrambles of small squares varying in gray level.

For simplicity, we will also assume that

$$Sal_\phi(\rho) = 2\|W_\phi F^T \rho\| \tag{21}$$

where F is the matrix whose columns are the three differential sensitivity functions f_1, f_2 and f_3, and

$$W_\phi = K\left(I + R_\phi\right) \tag{22}$$

for I the 3×3 identity matrix and R_ϕ the 3×3 diagonal matrix such that $R_\phi(k,k) = |f^T_k \phi|$, and the positive scalar K is chosen so that the sum of the entries in W_ϕ is equal to 3. In this choice of W_ϕ, we are imagining that the observer can partially tune his/her salience computation to the specific discrimination task at hand by giving higher weight to those field-capture channels most sensitive to ϕ (i.e., those field-capture channels for which $R_\phi(k,k) = |f^T_k \phi|$ is large). However, as reflected by the contribution of the identity matrix to W_ϕ, we are also imagining that the

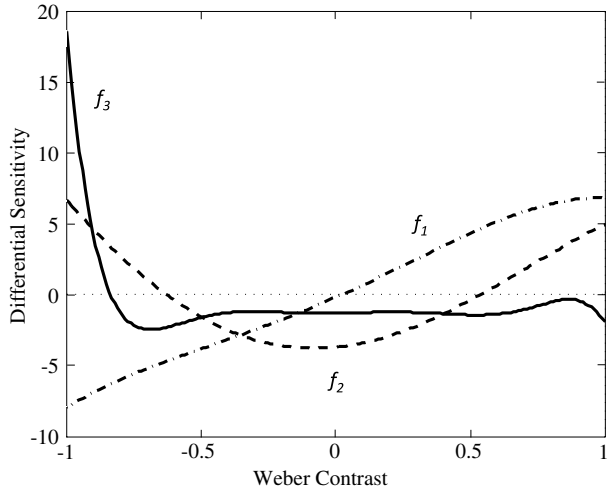

Fig. 6. The differential sensitivity functions of three hypothetical field-capture channels enabling discrimination of Weber contrast scrambles. These field-capture channels are roughly consistent with previous findings Chubb et al. (1994, 2004, 2007).

observer's flexibility in tuning his/her salience computation is limited. Also note that Eq. (21) results from setting $H_\phi(x) = x^{1/2}$, and $G_{\phi,k}(x) = (2W_\phi(k,k)x)^2$ in Eq. (6). Finally, we assume probability correct in our 4-option forced-choice orientation discrimination task is given by Eq. (7) with $\gamma = 2.5$, approximately in line with the results of Chubb et al. (2004).

5.2. Selecting the micropattern set

In this example, the domain of physical variation being investigated is continuous. The first question one must answer with this sort of physical domain is: How finely should the micropattern set Ω sample the domain? (or equivalently, how large should we take N_Ω?) This decision should be made carefully for several reasons.

1. There are at least two reasons to make N_Ω moderately large:
   (a) One needs to be confident that N_Ω is greater than the number N_{FCC} of field-capture channels differentially sensitive to Ω scrambles. Otherwise, it will be impossible either to derive a basis spanning the Ω differential sensitivity space or to determine its dimension.
   (b) If one or more field-capture channels has sensitivity that varies rapidly as a function of v, then if N_Ω is too small, one may fail to detect those variations.
2. On the other hand, there are also important reasons to keep N_Ω small:
   (a) If N_Ω is small, then scrambles can have higher concentrations of the different kinds of micropatterns in Ω, and one can experimentally vary these concentrations over a larger range. As this suggests, the smaller one makes Ω, the easier it tends to be to generate scrambles that are strongly discriminable, and the easier it is to measure sensitivities to those scrambles.
   (b) The number of trials required to estimate the expansion of a given seed using the method described above is roughly proportional to N_Ω, so the smaller one keeps N_Ω, the easier it is to derive a basis of the Ω sensitivity space.

We have found that taking N_Ω around 9 often works well. However, the particular choice will depend on factors specific to the domain of physical variation under study. In our simulations, $N_\Omega = 9$; specifically, we let our micropattern set Ω contain nine small squares with Weber contrasts -1.0, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1.0. For convenience, we identify the different micropatterns with their Weber contrasts.

5.3. Strategic considerations in choosing the first seed

In this section we use the seed-expansion method outlined in Section 4 to estimate a basis of the Ω sensitivity space spanned by the differential sensitivity functions shown in Fig. 6. To derive a basis of the Ω differential sensitivity space we will need to expand at least three different seeds, ϕ_1, ϕ_2, ϕ_3. We should find that the space of perturbations orthogonal to all of Δ_{ϕ_1}, Δ_{ϕ_2} and Δ_{ϕ_3} affords very weak discrimination (if any).

We shall choose our first seed from among the discrete domain Legendre polynomials, $\lambda_1, \lambda_2, \ldots, \lambda_{N_\Omega - 1}$ shown in Fig. 7. These perturbations are derived by applying Gram–Schmidt orthonormalization to the vectors $v_k(\omega) = \omega^k$, for $k = 0, 1, \ldots, N_\Omega - 1$. (We throw out $\lambda_0(\omega)$, which is the constant function $= \frac{1}{\sqrt{N_\Omega}}$ and therefore not a perturbation.) For any k between 1 and $N_\Omega - 1$, the perturbations $\lambda_1, \lambda_2, \ldots, \lambda_k$ span the space of all polynomial perturbations of order less than or equal to k.

We expect that human vision will have one or more field-capture channels providing broad, well-graded sensitivity to any linear domain of physical variation such as Weber contrast. Often the more interesting question is whether there exist any field-capture channels that confer "special" sensitivity showing some form of abrupt variation (e.g., sharp tuning to a particular contrast). The Legendre polynomials are useful for quickly discovering such forms of sensitivity if they exist. In practice, it suffices to glance at displays like those in Fig. 4 that use scrambles whose histograms are maximally (and oppositely) perturbed by the different Legendre polynomials. If one finds a high order Legendre λ_k that is easily discriminable, this reveals immediately that human vision has one or more field-capture channels with some form of abrupt variation in sensitivity. To discover the form of this sensitivity all one needs to do is use λ_k as a seed perturbation and extract the corresponding expansion.[4] For this reason we typically take ϕ_1, our first seed, to be the highest order Legendre polynomial that presents itself as obviously discriminable. In the current example, this is λ_4. (Scrambles with histograms $U + \lambda_4$ and $U - \lambda_4$ are shown in Fig. 3h and i.)[5]

5.3.1. Expanding the first seed

It is crucial to use a set of perturbations ρ that are all highly correlated (e.g., with correlation greater than around 0.9 with the seed).

It is also crucial that the ρ's one tests span the space of all perturbations. In this section, we will describe a specific approach that achieves both of these aims. We have used this method in a number of experiments, and it seems to work well.

1. We begin by selecting an orthonormal basis $b_1, b_2, \ldots, b_{N_\Omega}$ of perturbations that has b_1 equal to the seed. In the current case we let

$$\begin{aligned} b_1 &= \phi_1 = \lambda_4 \\ b_2 &= \lambda_1 \\ b_3 &= \lambda_2 \\ b_4 &= \lambda_3 \\ b_5 &= \lambda_5 \end{aligned} \tag{23}$$

[4] The experiments reported in Chubb et al. (1994) and Chubb et al. (2004) used to document the existence of the blackshot mechanism amount to an application of the seed expansion method. In these studies, Ω comprises the 17 gray levels $-1, -0.125, \ldots, 1$, the seed is λ_3, and the expansion is the projection of the blackshot sensitivity function into the space spanned by $\lambda_1, \lambda_2, \ldots, \lambda_7$.

[5] The maximal form of λ_4 has amplitude $A_{max} = 0.237$, and $Sal(A_{max}\lambda_4) = 2A_{max}|W_{\lambda_4}F^T\lambda_4| = 2.13$. In addition, $\Psi(2.13) = 1.00$ for Ψ given in Eq. (7) with $\gamma = 2.5$. The corresponding computation for λ_5 yields probability correct of 0.92, and the probabilities correct are lower for λ_6, λ_7 and λ_8.

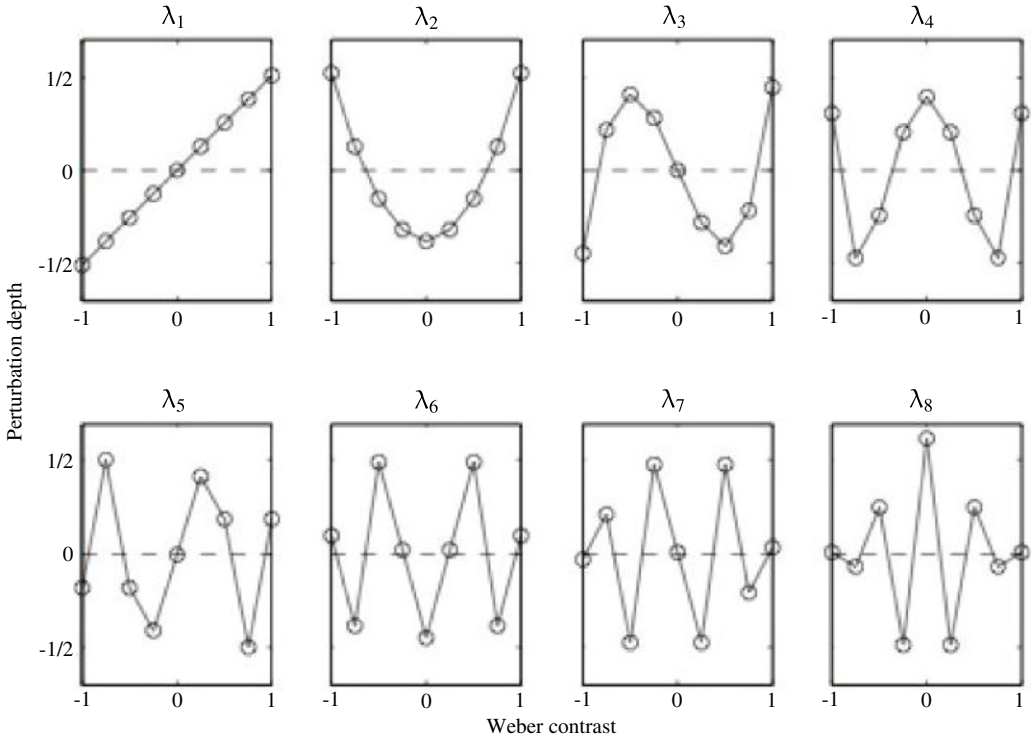

Fig. 7. Discrete domain Legendre polynomials of order 1–8. They form an orthogonal basis of the space of all perturbations of the nine Weber contrasts.

$b_6 = \lambda_6$

$b_7 = \lambda_7$

$b_8 = \lambda_8$.

It should be noted, however, that any choice of $b_2, \ldots, b_8$ will suffice as long as they are orthonormal and $b_1 = \phi_1$.

2. Then for $\epsilon = \frac{1}{3}$, we construct the perturbations

$$\eta_k^+ = \frac{b_1 + \epsilon b_k}{\|b_1 + \epsilon b_k\|} \tag{24}$$

and

$$\eta_k^- = \frac{b_1 - \epsilon b_k}{\|b_1 - \epsilon b_k\|} \tag{25}$$

for $k = 2, 3, \ldots, 8$.

3. Note that each of the perturbations $\rho = b_1$, as well as $\rho = \eta_k^+$ and $\rho = \eta_k^-$ for $k = 2, 3, \ldots, 8$ is normalized. For each of these perturbations ρ we collect psychometric data testing performance at discriminating $A\rho$ for various amplitudes A. We will use 15 interleaved staircases to collect this data, one for each perturbation. Specifically, the staircase for a given perturbation ρ is structured as follows:
   (a) For A_{max} the scalar for which the maximum absolute value of $A_{max}\rho$ is equal to $\frac{1}{N_\Omega}$ ($\frac{1}{9}$ in this example), we take as our range of possible histogram amplitudes A, the 30 values $\frac{A_{max}}{30}, \frac{2A_{max}}{30}, \ldots, A_{max}$.
   (b) Then we start the staircase at amplitude $A = \frac{A_{max}}{2}$ and proceed to observe 250 staircase trials for each of $\rho = b_1$, η_k^+ and η_k^-, $k = 2, 3, \ldots, 8$ (3750 trials total). In each staircase, we increment A whenever either
      i. fewer than two trials have been observed, or
      ii. one or both of the last two trials yielded an incorrect response,

   and otherwise we decrement A (i.e., whenever the previous two trials both yielded correct responses). (Staircases that use this "2-up-1-down" update rule concentrate observations around perturbation amplitudes that yield performance in the neighborhood of 71% correct.)
4. Next we use a Bayesian estimation procedure to derive a sample from the joint posterior density characterizing the parameters required to fit all of the psychometric data. These parameters are
   (a) γ from Eq. (7),
   (b) and the expansion Δ_{ϕ_1} (whose values are constrained to sum to 0).

   The estimation method uses Markov Chain Monte Carlo (MCMC) simulation. We refer the reader to Appendix for details.

The estimated basis function $f_{\phi_1} = \Delta_{\phi_1}$ is plotted in Fig. 8. Error bars are 95% Bayesian credible intervals.[6] The dashed line gives the actual value of f_{ϕ_1}. The dotted–dashed line gives the part of f_{ϕ_1} contributed by the seed perturbation λ_4; i.e., this is the projection of f_{ϕ_1} into the one-dimensional space spanned by λ_4. Note that λ_4 itself contributes relatively little to f_{ϕ_1}.

5.3.2. Testing the local linearity assumption

Note that we have collected psychometric data for varying amplitudes of each of the 15 normalized perturbations $\rho_1 = b_1$, as well as $\rho_{2k} = \eta_k^+$ and $\rho_{2k+1} = \eta_k^-$ for $k = 1, 2, \ldots, 7$. These conditions are redundant with respect to the goal of estimating Δ_{ϕ_1} and thus enable us to test the basic assumption of our method (expressed in Eq. (19)) that $Sal_{\phi_1}(\rho)$ can be approximated by the linear transformation.

It is convenient to use a likelihood ratio test for this purpose. This general method can be used to compare two nested models fit using a maximum likelihood criterion. In particular, suppose the free parameters in $Model_{\text{Constrained}}$ comprise a subset of those in $Model_{\text{General}}$. Let $\tilde{\Lambda}_{\text{General}}$ be the maximum value of the likelihood function for $Model_{\text{General}}$ across all possible values of the model

[6] In this simulation, all threshold estimates are medians of estimated Bayesian posterior distributions, and bracketing markers give the 2.5 and 97.5 percentiles of these same distributions.

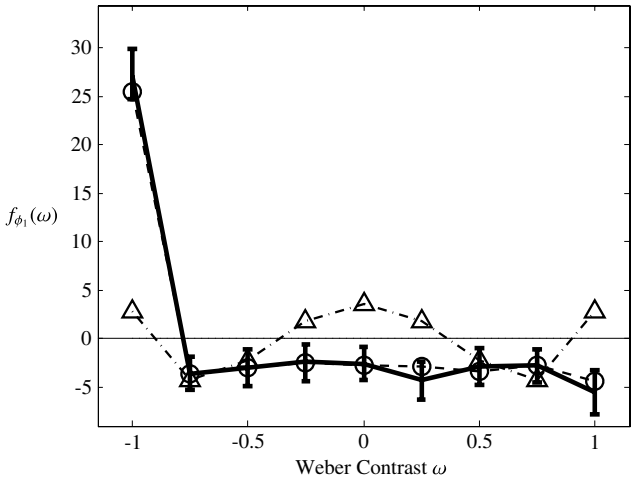

Fig. 8. The first basis element, f_{ϕ_1}. This is the expansion of the seed perturbation $\phi_1 = \lambda_4$ (Fig. 7). Error bars are 95% Bayesian credible intervals. The dashed line gives the true value of f_{ϕ_1}. The dotted–dashed line gives the contribution to f_{ϕ_1} of the seed perturbation ϕ_1.

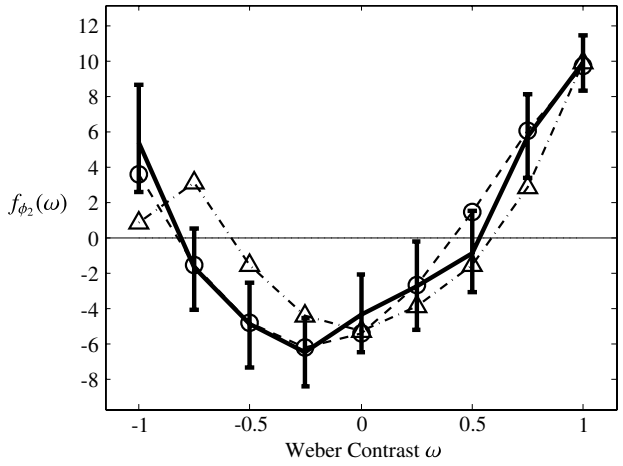

Fig. 9. *The second basis element,* f_{ϕ_2}. This is the expansion of the seed perturbation ϕ_2 derived by orthonormalizing λ_2 (7) with respect to the first basis element f_{ϕ_1} shown in Fig. 8. Error bars are 95% Bayesian credible intervals. The dashed line gives the true value of f_{ϕ_2}. The dotted–dashed line gives the contribution to f_{ϕ_2} of the seed perturbation ϕ_2.

parameters, and let $\tilde{\Lambda}_{\text{Constrained}}$ be the maximum value of the likelihood function for $Model_{\text{Constrained}}$. As is well known (see, e.g., Hoel, Port, & Stone, 1971), if $Model_{\text{Constrained}}$ reflects the true state of nature (this is the null hypothesis), then the statistic

$$X = -2 \ln \left(\frac{\tilde{\Lambda}_{\text{Constrained}}}{\tilde{\Lambda}_{\text{General}}} \right) \tag{26}$$

is asymptotically distributed as chi-square with degrees of freedom equal to the difference between the number of degrees of freedom in $Model_{\text{General}}$ and the number in $Model_{\text{Constrained}}$.

In the current situation the likelihood $\tilde{\Lambda}_{\text{Constrained}}$ is obtained by maximizing Λ_ϕ in Eq. (20) across all possible values of Δ_ϕ and γ. The number of free parameters in this model is 9. (Although the function Δ_ϕ has nine values, it only has eight degrees of freedom because it is constrained to sum to 0.)

For $Model_{\text{General}}$ we fix γ across the fifteen perturbations ρ_k, $k = 1, 2, \ldots, 15$ and introduce separate scaling parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$ to fit the data from each staircase separately. Let $C_k(A)$ ($I_k(A)$) be the number of trials on which the perturbation presented was $A\rho_k$ and the participant responded correctly (incorrectly). Then

$$\Lambda_{\text{General}}(\alpha_1, \alpha_2, \ldots, \alpha_{15}, \gamma) = \prod_{k=1}^{15} \prod_{\text{norms } A} \Psi\left(\frac{A}{\alpha_k}\right)^{C_k(A)} \times \left(1 - \Psi\left(\frac{A}{\alpha_k}\right)\right)^{I_k(A)}, \tag{27}$$

and $\tilde{\Lambda}_{\text{General}}$ is the maximum value taken by Λ_{General} across all possible values of its 16 parameters.

To find the values of each of $\tilde{\Lambda}_{\text{General}}$ and $\tilde{\Lambda}_{\text{Constrained}}$ it is necessary to conduct a search using some optimization program. As usual, one must be careful to check that each of these searches has found the global maximum of the function across its entire domain. This is likely to require starting the search at a number of different locations in the parameter space.

In the current instance, we find that $\ln(\tilde{\Lambda}_{\text{General}}) = -2550.103$, $\ln(\tilde{\Lambda}_{\text{Constrained}}) = -2552.971$. Thus the statistic X of Eq. (26) takes the value $2 \times (2552.971 - 2550.103) = 5.7360$. Under the null hypothesis, the probability of obtaining a value this large or greater is $1 - F_{\chi^2,7}(5.7360) = 0.571$, where $F_{\chi^2,7}$ is the chi-square cumulative distribution function with 7 degrees of freedom. Thus, our current data give us no reason to doubt our assumption of linearity.

5.3.3. Deriving the second basis function

To find our second seed we project the Legendre polynomials $\lambda_1, \ldots, \lambda_8$ into the space of perturbations orthogonal to f_{ϕ_1} to obtain normalized perturbations $g_1, \ldots, g_8$ and seek the highest k for which the maximal form of g_k yields near perfect discrimination. In the current example, this is g_2 (whose maximal form yields probability 0.998 correct discrimination). We take the normalized perturbation

$$b_1 = \phi_2 = g_2 \tag{28}$$

as our second seed. This function (scaled to reflect its contribution to the second expansion) is shown by the dotted–dashed line with triangular markers in Fig. 9. We then fill out the basis $b_2, \ldots, b_8$ with an arbitrary set of orthonormal functions all orthogonal to b_1.[7]

From here, the procedure for measuring the second expansion is identical to the procedure we used to measure the first expansion. For details, see Steps 2–4 of Section 5.3.1.

The estimate of the expansion $f_{\phi_2} = \Delta_{\phi_2}$ of ϕ_2 (obtained by our simulated experiment) is given by the black line in Fig. 9. The error bars are 95% Bayesian credible intervals, and the true value of f_{ϕ_2} is shown by the dashed line with circular markers.

5.3.4. Deriving the third basis function

The third seed ϕ_3 is the normalized projection of λ_1 into the space orthogonal to f_{ϕ_1} and f_{ϕ_2}. The maximal form of this perturbation yields virtually perfect performance (probability correct $= 0.999$). As before, we set b_1 equal the seed to be expanded (ϕ_3). Then we choose an arbitrary set of other perturbations $b_2, \ldots, b_8$ so as to make the set of b_k's orthonormal. The procedure for deriving the expansion f_{ϕ_3} is identical to the procedure used for each of the first two expansions.

The estimate derived of the third expansion f_{ϕ_3} is shown by the black line in Fig. 10. Error bars give 95% Bayesian credible intervals. The dotted–dashed line with triangular markers shows the contribution of the seed perturbation ϕ_3 to the expansion f_{ϕ_3}, and the dashed line with circular markers shows the true value of f_{ϕ_3}.

[7] Specifically, we use the Matlab command, "Others = null([ones(1, 9); ones (8, 1) * b1(:)']);" This produces an orthonormal basis of the space orthogonal to both the constant function U and the seed perturbation, b_1.

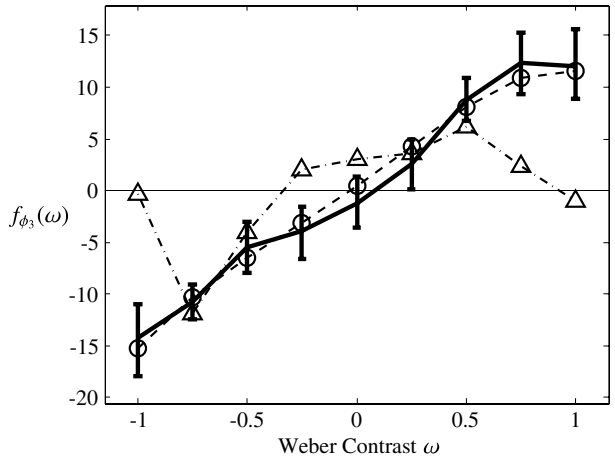

Fig. 10. *The third basis element, f_{ϕ_3}.* This is the expansion of the seed perturbation ϕ_3 derived by orthonormalizing λ_1 (7) with respect to the first and second basis elements f_{ϕ_1} (Fig. 8) and f_{ϕ_2} (Fig. 9). Error bars are 95% Bayesian credible intervals. The dashed line gives the true value of f_{ϕ_3}. The dotted–dashed line gives the contribution to f_{ϕ_3} of the seed perturbation ϕ_3.

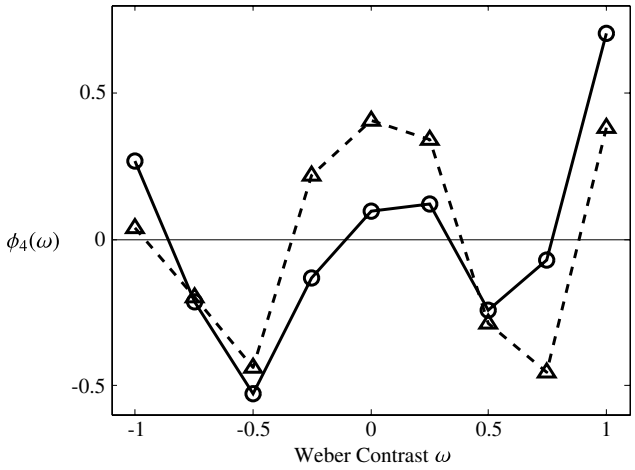

Fig. 11. The fourth seed. The normalized perturbation g shown by the dashed line is derived by orthonormalizing λ_1 (See Fig. 7) with respect to the first three expansions. The maximal form of this perturbation yields discrimination performance that is too poor to allow g to be used as a seed in the seed expansion method but significantly greater than chance, showing that there exist field-capture channels that are differentially sensitive to g. To explore the sensitivity to g, we take a mixture of g with enough f_{ϕ_2} to produce a normalized perturbation ϕ_4 whose maximal form yields near perfect discrimination performance. Specifically, ϕ_4 is derived by normalizing the perturbation $\frac{f_{\phi_2}}{\|f_{\phi_2}\|} + g$. The maximal form of this perturbation yields proportion correct 0.973.

5.3.5. Testing the residual space of perturbations

After each iteration of the seed-expansion method, one must check that there still exists some perturbation orthogonal to the previously obtained basis functions that affords discrimination above chance. We have not emphasized this stage in previous iterations because the maximal form of each of the perturbations used as a seed yields obvious discrimination. For ρ equal to the maximal form of ϕ_k, $k = 1, 2, 3$, a glance at a square wave modulating between scrambles with histograms $U + \rho$ and $U - \rho$ would immediately reveal that discrimination will be superthreshold in our standard task.

However, the space of perturbations orthogonal to f_{ϕ_1}, f_{ϕ_2} and f_{ϕ_3} yields no obvious candidates for a fourth seed. A good guess at such a fourth seed can usually be derived by projecting λ_1 (or some simple power function) into the space of perturbations orthogonal to all of the previous expansions. In the current case, this yields the normalized perturbation g shown by the dashed line in Fig. 11. The maximal from of the perturbation g yields discrimination performance that is subthreshold but significantly greater than chance (proportion correct $= 0.478$). On the one hand, the fact that the maximal form of g yields greater-than-chance performance indicates that there exist field-capture channels with some sensitivity to g. On the other hand, g cannot play the role of the seed perturbation in the seed expansion method, because the level of performance supported by its maximal form is too low. In order to measure an expansion, we need a seed that supports near perfect discrimination. Thus, to investigate the residual sensitivity to g, we will need to mix g with some other perturbation that supports near perfect discrimination. In the current case, we normalize the perturbation $\frac{f_{\phi_2}}{\|f_{\phi_2}\|} + g$ to derive our fourth seed, ϕ_4. The maximal form of ϕ_4 yields proportion correct discrimination 0.973. The choice to use f_{ϕ_2} in this role is dictated by the fact (as revealed by a glance at Figs. 8–10) that the credible intervals are larger for our estimate of f_{ϕ_2} than for either of the other two expansions. Thus in using f_{ϕ_2} to enrich ϕ_4, we focus selectively on the region of Ω sensitivity space about which we have the poorest information.

We use the same method as in each of the three previous cases to derive the fourth expansion f_{ϕ_4}. That is, we select an arbitrary orthonormal basis $b_1, b_2, \ldots, b_8$ of perturbations with $b_1 = \phi_4$ and follow steps 2–4 of Section 5.3.1. The resulting expansion f_{ϕ_4} is shown in Fig. 12.

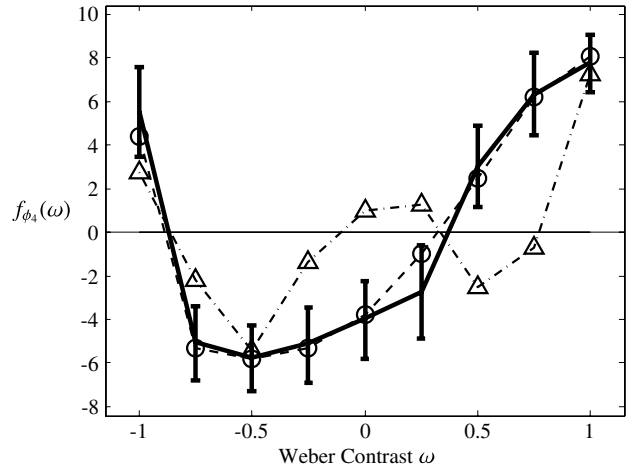

Fig. 12. *The fourth basis element, f_{ϕ_4}.* This is the expansion of the seed perturbation ϕ_4 derived by orthonormalizing λ_1 (7) with respect to the first and second basis elements f_{ϕ_1} (Fig. 8) and f_{ϕ_2} (Fig. 9). Error bars are 95% Bayesian credible intervals. The dashed line gives the true value of f_{ϕ_4}. The dotted–dashed line gives the contribution to f_{ϕ_4} of the seed perturbation ϕ_4. Note that even though there are only three field-capture channels in the example simulation, it is possible to derive the expansion of a fourth seed perturbation. Indeed, one can derive the expansions of as many different seed perturbations as one desires. However, all of these different expansions must reside (within measurement error) in the three-dimensional subspace spanned by the sensitivity functions plotted in Fig. 6.

We might reiterate this process one or even two more times. However, the fact that we were unable to find a perturbation orthogonal to all of f_{ϕ_1}, f_{ϕ_2}, and f_{ϕ_3} that supported strong discrimination performance suggests that the space orthogonal to $f_{\phi_1}, f_{\phi_2}, f_{\phi_3}$ and f_{ϕ_4} should be a subspace of the Ω null space. We can investigate this question more carefully by extracting the principal components of the expansions $f_{\phi_1}, f_{\phi_2}, f_{\phi_3}$ and f_{ϕ_4} as described in the next section.

5.4. Extracting a basis

To derive a basis of the sensitivity space, we proceed to use singular value decomposition. In deriving credible intervals for

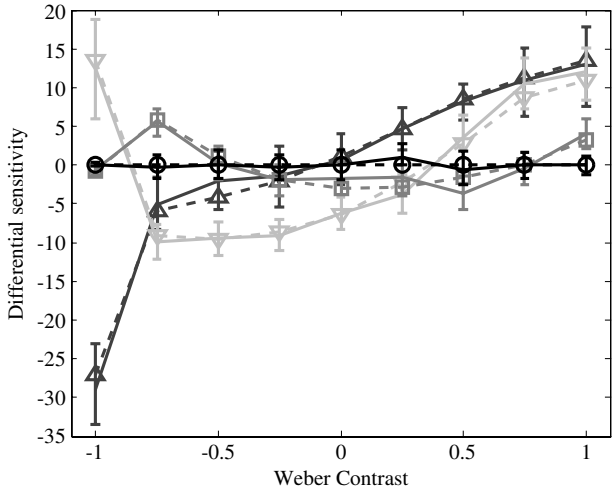

Fig. 13. Means and associated 95% Bayesian credible intervals for the principal components of the 9 × 4 matrix whose columns are the four expansions. Each principal component has been scaled by its corresponding eigenvalue. The dashed lines give the true values of the four functions. Note that the estimate of the fourth principal component (plotted with circles) does not differ significantly from 0 whereas each of the other three estimates does. This supports the claim that the Ω sensitivity space is the three-dimensional space spanned by estimates of the three curves marked with tokens other than circles.

each of the four expansions (plotted in Figs. 8–10 and 12) we used MCMC simulation to produce 40,000 samples drawn from the posterior distribution characterizing each expansion. For $k = 1, 2, \ldots, 40{,}000$, we apply singular value decomposition to the matrix M_k whose columns are the kth samples of expansions 1 through 4. For each k, this yields the (unique) decomposition

$$M_k = S_k V_k D_k^T \tag{29}$$

where

1. S_k is the 9 × 4 matrix whose columns are the orthonormal "principal components" of M_k,
2. V_k is the diagonal matrix whose ith diagonal element gives the eigenvalue associated with the ith principal component, and
3. D_k is the 4 × 4 orthonormal matrix whose jth row gives the "loadings" of the four principal components in the jth column vector of M_k; this matrix is irrelevant for our purposes.

Fig. 13 plots the means and associated 95% Bayesian credible intervals for the four columns of the matrix $S_k V_k$, $k = 1, 2, \ldots, 40{,}000$. The ith column ($i = 1, 2, 3, 4$) of this matrix gives the ith principal component scaled by its eigenvalue. The dashed lines give the true values of the eigenvalue-scaled principal components, which we will simply call "principal components" from now on.

Note first that the estimated principal components fit the real ones within measurement error. Crucially, the fourth principal component does not differ significantly from 0 whereas the other three all do. The analysis we have performed has thus confirmed that the Ω differential sensitivity space is three dimensional. Specifically, it is the space spanned by the three curves in Fig. 13 with non-circular markers.

6. Discussion

6.1. Moving beyond scrambles

In the past decade several very effective texture synthesis methods have been developed (e.g., Portilla & Simoncelli, 2000; Zhu et al., 1999, 1996). Given a sample T of some target texture, each of these methods uses a stochastic process to derive a new random image R equated to the original texture in a specified set P of statistics. If the method used to derive R draws uniformly from the set of all images equated to T in statistics P, it is said to be a "maximum entropy" procedure. Maximum entropy methods are preferred above others because only these methods yield textures whose properties are purely entailed by the constraint that R match T in statistics P. Such procedures have been explored by Zhu et al. (1999, 1996). By contrast, the method of Portilla and Simoncelli (2000) is not a maximum entropy procedure; however, it plausibly approximates this ideal and is much more efficient than available maximum entropy methods. Portilla and Simoncelli (2000) start with an $N_{rows} \times N_{cols}$ image whose pixels are initialized to jointly independent, standard normal random variables; the procedure then iteratively nudges that image through $N_{rows}N_{cols}$-dimensional space so as to arrive at an image equated to T in statistics P. Portilla and Simoncelli (2000) have worked hard to derive a fairly trim set P of statistics that seems to be sufficient to describe human preattentive sensitivity. That is, for all textures T that they have tested, the synthesized images R produced to match T in statistics P end up being preattentively indiscriminable from T, or nearly so. However, despite the efforts of Portilla and Simoncelli (2000) to trim the fat from the set P, there is no guarantee that P is minimal. On the contrary, it seems highly likely that P is far from minimal. The challenge, then, is to distill the sufficient set P into a core set of necessary statistics.

Although we have focused on discrimination of scrambles in the current paper, the formalism we have developed transfers directly to investigating discrimination of synthetic textures. We will sketch how this works. First, however, to build a bridge from scrambles to synthetic textures, we need to foreground one aspect of the scramble experiments that we have not previously emphasized. In all of the scramble experiments we have been discussing, the scramble with histogram U plays the role of our "base" texture: all pairs of scrambles to be discriminated in the experiments we have described deviate symmetrically from the scramble with histogram U. That is, on every trial, the observer is asked to discriminate a scramble with some histogram $U + \rho$ from the scramble with complementary histogram $U - \rho$. It should be noted, however, that one might well use some other histogram Y instead of U to define the "base" texture in a scramble experiment; and if one did, the results might change: one might obtain a different sensitivity space using Y than using U. Implicit in the logic of the seed expansion method is the understanding that one's results are conditioned by the base texture one is using. In effect, what we derive in an iterated seed expansion experiment is the human sensitivity space for perturbations *away from the base texture* used in the experiment.

Let us now imagine how the apparatus we have developed might be used to study discrimination of synthetic textures. For concreteness we imagine an experiment that makes use of the Portilla–Simoncelli algorithm. Suppose the number of Portilla–Simoncelli (P–S) statistics is N. Just as any Ω-scramble is defined by its histogram, any P–S texture is defined by the N values it assigns statistics P. Analogous to the histogram U that has defined the base scramble in all the experiments we have considered, we first need to choose some specific $V_0 \in \mathbb{R}^N$ to define the "base" P–S texture in our experiment. In this context, any vector $\rho \in \mathbb{R}^N$ can play the role of a perturbation, and one can measure the discriminability of the textures with statistics $V_0 + \rho$ vs. $V_0 - \rho$. We will use the same shorthand as we did for scrambles: we will say the perturbation "ρ is discriminable with probability p" if the participant achieves success rate p in discriminating the synthetic textures with statistics $V_0 + \rho$ vs. $V_0 - \rho$; and if p is high, we say "ρ is easily discriminable".

Extending the analogy to a full-blown seed expansion experiment, let $B = \{b_1, b_2, \ldots, b_M\}$ be a set of orthonormal vectors in $\mathbb{R}^N$ and let $Span(B)$ be the space of all linear combinations of B.

The question we address in this hypothetical experiment is, how many field-capture channels are differentially sensitive to the texture variations that can be produced by linearly combining the perturbations in B in the neighborhood of V_0, and what sorts of sensitivity do they confer?

The crucial assumption we need to make is that the mean activation produced in any given field-capture channel by a texture varies slowly as a function of the P–S statistics of the texture; more specifically, we must assume that *the mean response of any field-capture channel is approximately a linear function of P–S statistics within any neighborhood small enough that the corresponding textures support only subperfect discrimination.* Formally: if two textures with P–S statistics V_1 and V_2 support subperfect discrimination in a standard task,[8] then we assume that the following condition holds to a good approximation for any field-capture channel C: for some $f_C \in \mathbb{R}^N$, the mean activation produced in C by the texture with P–S statistics equal to $AV_1 + (1 - A)V_2$ is equal to $Af_C^T V_1 + (1 - A)f_C^T V_2$ for A ranging from 0 to 1.

If this "local linearity" condition holds, then the experimental methods presented in this paper can be used to (1) determine the number of field-capture channels whose differential sensitivity functions have non-zero projections into the space spanned by $B = \{b_1, b_2, \ldots, b_M\}$, and (2) derive a basis of the sensitivity space $S_{V_0,B}$ spanned by those projections. The procedure is exactly analogous to the one described in Section 4. One starts by finding a normalized perturbation $\phi_1 \in Span(B)$ that can be scaled so as to be easily discriminable. Then one extracts the expansion $f_{\phi_1} \in Span(B)$ of ϕ_1. Under the local linearity assumption, f_{ϕ_1} is an element of $S_{V_0,B}$. Next, one selects another seed perturbation $\phi_2 \in Span(B)$ such that (1) ϕ_2 is orthonormal to f_{ϕ_1}, and (2) ϕ_2 can be scaled to yield near perfect discrimination; then one extracts the expansion $f_{\phi_2} \in Span(B)$ of ϕ_2. The same logic applies to support the conclusion that $f_{\phi_2} \in S_{V_0,B}$. One iterates this procedure as many times as necessary to a obtain a matrix $F_{V_0,B}$ whose columns (the successive expansions) can be confidently taken to span the sensitivity space $S_{V_0,B}$. The principal components of $F_{V_0,B}$ whose eigenvalues are significantly greater than 0 provide an orthonormal basis of the sensitivity space $S_{V_0,B}$.

Why is this important? First, the dimension $D_{V_0,B}$ of $S_{V_0,B}$ places a lower bound on the number of field-capture channels that are differentially sensitive to P–S statistic variations in the neighborhood of V_0. Moreover, if the perturbations $b_1, b_2, \ldots, b_M$ are chosen judiciously, then one may have reason to believe that $D_{V_0,B}$ is actually equal to the total number of field-capture channels that are differentially sensitive to V_0. Second, the specific patterns of sensitivity reflected by $S_{V_0,B}$ are likely to provide important insights into the computations performed by the field-capture channels resident in human vision.

By reiterating this iterated seed expansion procedure across a number of different base textures (characterized by different vectors of P–S statistics) one might hope to derive data sufficient to enable a complete theory of human preattentive visual sensitivity.

6.2. Summary

A basic question in visual perception is: which physical variations in the pattern of light presented to the retina are spontaneously visible? At present, this question has only been partially answered. In this paper, we have described and demonstrated a method that can be used to answer this question for the space of textures that can be made by taking random arrangements micropatterns drawn in different proportions from a given set Ω. Such textures are called scrambles. If human vision has field-capture channels $M_1, M_2, \ldots, M_N$ that are differentially sensitive to scrambles, then the methods we have described can be used to derive a basis of the space spanned by the functions $f_1, f_2, \ldots, f_N$ that give the sensitivities of these field-capture channels to the different micropatterns in Ω. Basis elements are extracted one at a time using a new method called "seed expansion" (see Section 4). We used the method in a simulated experiment, showing how to derive a complete basis of the differential sensitivity space of a micropattern set Ω by iterative seed-expansion. By applying the methods we have described across a range of different micropattern sets Ω it should be possible to derive the field-capture channels that observers have available to process these patterns.

Finally, we have described how the methods developed here might be applied to analyze human sensitivity to variations between textures other than scrambles. This is a vast and exciting domain that remains to be explored.

Acknowledgment

This work was supported by NSF BCS-0843897.

Appendix. Concerning the Bayesian analysis of psychometric data

In this appendix, we describe the details of the Bayesian method used to estimate the joint posterior density of the vector comprising the parameters used to fit the psychometric data collected in using the seed expansion method. This may be of use to readers seeking an introduction to Bayesian model estimation in psychophysical applications.

The estimation method uses Markov Chain Monte Carlo (MCMC) simulation. For simplicity, we will use uniform prior distributions on all parameters. In any MCMC process using uniform priors, one starts with some arbitrary guess V at the parameter vector (which will ultimately be thrown away) and sets ${}_1S = V$; then one iterates the following steps some large number N of times. (We will use pre-subscripts to indicate sample number in the MCMC process and ordinary subscripts to indicate the coordinate within a given sample.) In using this method to estimate the expansion f_{ϕ_k} and the Weibull function exponent γ, a given parameter vector comprised guesses at the values $f_{\phi_k}(\omega)$, $\omega = -1.0, -0.75, \ldots, 1.0$ and γ. Each parameter vector was also constrained to have the values f_{ϕ_k} sum to 0. For $n = 2, 3, \ldots, N$, randomly select a candidate parameter vector C in the neighborhood of ${}_{n-1}S$. Then[9] for

$$ {}_nR = \frac{\Lambda(C)}{\Lambda({}_{n-1}S)} \tag{31} $$

- if ${}_nR \geq 1$, set ${}_nS = C$;
- otherwise set

$$ {}_nS = \begin{cases} C & \text{with probability } {}_nR \\ {}_{n-1}S & \text{with probability } 1 - {}_nR. \end{cases} \tag{32} $$

In practice, to keep the computation within range of floating point representation, one never actually computes $\Lambda(C)$ or $\Lambda({}_{n-1}S)$; rather, one computes $LogL_C = ln(\Lambda(C))$ and $LogL_{{}_{n-1}S} = ln(\Lambda({}_{n-1}S))$, and then sets ${}_nR = exp\left(LogL_C - LogL_{{}_{n-1}S}\right)$.

The classical result (e.g., Hastings, 1970) is that in the limit as $N \to \infty$ this algorithm yields a sample from the posterior density.

[8] For example, a task in which the observer is presented with four texture patches presented in a square array and must pick the odd man out (the one with P–S statistics different from the other three).

[9] If the prior density f_{prior} were nonuniform, then we would have

$$ {}_nR = \frac{\Lambda(C)f_{prior}(C)}{\Lambda({}_{n-1}S)f_{prior}({}_{n-1}S)}. \tag{30} $$

Priors

The bounds of the uniform densities one uses to define the priors matter very little provided they are sufficiently inclusive so as not to cut off any part of the posterior density. In the current simulations, the prior density of $f_{\phi_k}(\omega)$ was taken to be uniform between -1000 and 1000 for all nine Weber contrasts $\omega = -1.0, -0.75, \ldots, 1.0$, and the prior density of γ was taken to be uniform on the interval from 0 to 1000. As candidate parameter vectors C were drawn, the program checked to make sure that each coordinate value C_k was within the upper and lower boundaries of its prior density.

Adaptive candidate selection

As noted above, on the nth iteration of the MCMC process, one randomly selects a candidate parameter vector C in the neighborhood of ${}_{n-1}S$. The window used to perform this sampling (i.e., how one defines the sampling neighborhood) dramatically influences the efficiency with which one can estimate the posterior joint density of the parameters. We adjust this sampling window adaptively after each 1000 iterations of the MCMC process. Specifically, let S_{last1000} be the matrix whose columns are the 1000 most recent parameter vectors added to the list by the MCMC process. We first subtract the mean of these 1000 parameter vectors from each vector in S_{last1000} to generate a matrix D_{last1000}. We use singular value decomposition to extract (1) the matrix Q whose columns are the (orthonormal) principal components of D_{last1000} as well as (2) the diagonal matrix V whose kth diagonal entry is eigenvalue of the kth column of Q. In each of the subsequent 1000 iterations of the MCMC process, we draw each successive candidate parameter vector ${}_kC$ by setting ${}_kC =_{k-1} S + QVX$ where X comprises a vector of independent normal random variables with mean 0 and standard deviation $\frac{1}{30}$. In essence, we use the last 1000 parameter vectors to approximate the posterior density as an elliptical cloud, and take steps scaled to the axes of this cloud. This method succeeds in achieving an MCMC process that moves efficiently to scribble in the joint posterior density.

Starting values, burn-in, and number of iterations

By default, our routine initializes the first 9 values of ${}_1S$ (corresponding to f_{ϕ_k}) to 0 and the last value of ${}_1S$ (corresponding to γ) to 2. Other starting points were tested; all yielded similar estimates of the posterior joint density. 50,000 iterations of the MCMC process were performed, the first 10,000 of these were used to allow the MCMC process to "burn in" (i.e., to find its way to the region of parameter space within which the joint posterior density assigns its mass), and the last 40,000 were taken as a representative sample of the posterior joint density characterizing the parameters f_{ϕ_k}, and γ.

References

Adelson, E. H., & Bergen, J. R. (1991). The plenoptic function and the elements of early vision. In M. S. Landy, & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 3–20). Cambridge, MA: MIT Press.

Beck, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science*, *154*, 538–540.

Beck, J. (1982). Textural segmentation. In J. Beck (Ed.), *Organization and representation in perception* (pp. 285–317). Hillsdale, NJ: Erlbaum.

Beck, J., Prazdny, K., & Rosenfeld, A. (1983). A theory of textural segmentation. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 1–38). Academic Press.

Beck, J., Sutter, A., & Ivry, R. (1987). Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing*, *37*, 299–325.

Bergen, J. R., & Adelson, E. H. (1988). Visual texture segmentation based on energy measures. *Journal of the Optical Society of America A*, *3*, 98–101.

Bergen, J. R., & Landy, M. S. (1991). The computational modeling of visual texture segregation. In M. S. Landy, & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 253–271). Cambridge, MA: MIT Press.

Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*, 55–73.

Caelli, T. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision*, *1*, 19–30.

Chubb, C. (1999). Texture-based methods for analyzing elementary visual substances. *Journal of Mathematical Psychology*, *43*, 539–567.

Chubb, C., Econopouly, J., & Landy, M. S. (1994). Histogram contrast analysis and the visual segregation of iid textures. *Journal of the Optical Society of America A*, *11*, 2350–2374.

Chubb, C., Landy, M. S., & Econopouly, J. (2004). A visual mechanism tuned to black. *Vision Research*, *44*, 3223–3232.

Chubb, C., & Nam, J.-H. (2000). The variance of high contrast texture is sensed using negative half-wave rectification. *Vision Research*, *40*, 1695–1709.

Chubb, C., Nam, J.-H., Bindman, D. R., & Sperling, G. (2007). The three dimensions of human visual sensitivity to first-order contrast statistics. *Vision Research*, *47*, 2237–2248.

Chubb, C., & Talevich, J. (2002). Attentional control of texture orientation judgments. *Vision Research*, *42*, 311–330.

Diaconis, P., & Freedman, D. (1981). On the statistics of vision: the julesz conjecture. *Journal of Mathematical Psychology*, *23*, 112–118.

Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminators. *Biological Cybernetics*, *61*, 103–113.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Graham, N. (1991). Complex channels, early local nonlinearities, and normalization in texture segregation. In M. S. Landy, & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 273–290). Cambridge, MA: MIT Press.

Graham, N., Beck, J., & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: effects of sign and amount of contrast. *Vision Research*, *32*, 719–743.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, *38*, 141–171.

Gurnsey, R., & Browse, R. A. (1989). Asymmetries in visual texture discrimination. *Spatial Vision*, *4*, 31–44.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97–109.

Hoel, P. G., Port, S. C., & Stone, C. J. (1971). *Introduction to statistical theory*. Boston, MA: Houghton-Mifflin.

Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions on Information Theory*, *IT-8*, 84–92.

Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American*, 34–43.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, *290*, 91–97.

Julesz, B., & Bergen, J. R. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal*, *62*, 1619–1645.

Julesz, B., Gilbert, E. N., Sheppand, L. A., & Frisch, H. L. (1973). Inability of humans to discriminate between visual textures that agree in second order statistics. *Perception*, *2*, 391–405.

Julesz, B., Gilbert, E. N., & Victor, J. D. (1978). Visual discrimination of textures with identical third order statistics. *Biological Cybernetics*, *31*, 137–149.

Knutsson, H., & Granlund, G. H. (1983). Texture analysis using two-dimensional quadrature filters. In *Proceedings of the IEEE computer society workshop on computer architecture for pattern analysis and image database management* (pp. 206–213). Silver Spring, MD: IEEE Computer Society.

Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, *31*, 679–691.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, *7*, 923–932.

Maxwell, J. C. (1855). Experiments on colour, as perceived by the eye with remarks on colour-blindness. *Transactions of the Royal Society of Edinburgh*, *XXI, Part II*, 275–298.

Nam, J.-H., & Chubb, C. (2000). Texture luminance judgments are approximately veridical. *Vision Research*, *40*, 1677–1694.

Pollack, I. (1971a). Detection of one-, two-, and three-dimensional Markov constraints in visual displays. *Acta Psychologica*, *35*, 219–232.

Pollack, I. (1971b). Perception of two-dimensional markov constraints within visual displays. *Perception and Psychophysics*, *9*, 461–464.

Pollack, I. (1972). Visual discrimination thresholds for one- and two-dimensional markov spatial constraints. *Perception and Psychophysics*, *12*, 161–167.

Pollack, I. (1973). Discrimination of third-order markov constraints within visual displays. *Perception and Psychophysics*, *13*, 276–280.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–71.

Robson, J. G. (1980). Neural images: The physiological basis of spatial vision. In C. S. Harris (Ed.), *Visual coding and adaptability* (pp. 177–214). Hillsdale, NJ: Erlbaum.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Wilson, H.R. (1993). Nonlinear processes in visual pattern discrimination. In *Proceedings of the national academy of sciences*, USA 90, pp. 9785–9790.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how they do it. *Nature Reviews Neuroscience*, *5*, 1–7.

Zhu, S.C., Liu, X., & Wu, Y.N. (1999). Exploring the julesz ensemble by efficient Markov chain Monte Carlo. In *IEEE workshop on statistical and computational theories of vision*. pp. http://www.cls.ohio–state.edu/?szhu/SCTV99.html.

Zhu, S. C., Wu, Y. N., & Mumford, D. (1996). Filters, random fields and maximum entropy (frame)—towards the unified theory for texture modeling. *IEEE Conference on Computer Vision and Pattern Recognition*, 693–696.