

Learning optimal integration of arbitrary features in a perceptual discrimination task

Melchi M. Michel

Center for Perceptual Systems and Department of Psychology,
University of Texas at Austin, Austin, TX, USA



Robert A. Jacobs

Department of Brain and Cognitive Sciences,
University of Rochester, Rochester, NY, USA



A number of studies have demonstrated that people often integrate information from multiple perceptual cues in a statistically optimal manner when judging properties of surfaces in a scene. For example, subjects typically weight the information based on each cue to a degree that is inversely proportional to the variance of the distribution of a scene property given a cue's value. We wanted to determine whether subjects similarly use information about the reliabilities of arbitrary low-level visual features when making image-based discriminations, as in visual texture discrimination. To investigate this question, we developed a modification of the classification image technique and conducted two experiments that explored subjects' discrimination strategies using this improved technique. We created a basis set consisting of 20 low-level features and created stimuli by linearly combining the basis vectors. Subjects were trained to discriminate between two prototype signals corrupted with Gaussian feature noise. When we analyzed subjects' classification images over time, we found that they modified their decision strategies in a manner consistent with *optimal feature integration*, giving greater weight to reliable features and less weight to unreliable features. We conclude that optimal integration is not a characteristic specific to conventional visual cues or to judgments involving three-dimensional scene properties. Rather, just as researchers have previously demonstrated that people are sensitive to the reliabilities of conventionally defined cues when judging the depth or slant of a surface, we demonstrate that they are likewise sensitive to the reliabilities of arbitrary low-level features when making image-based discriminations.

Keywords: classification images, perceptual template, cue integration, perceptual learning

Citation: Michel, M. M., & Jacobs, R. A. (2008). Learning optimal integration of arbitrary features in a perceptual discrimination task. *Journal of Vision*, 8(2):3, 1–16, <http://journalofvision.org/8/2/3/>, doi:10.1167/8.2.3.

Introduction

Vision researchers have long realized that adult observers can be trained to improve their performance in simple perceptual tasks. Improvements with practice in visual acuity, hue perception, and velocity discrimination, for example, have been documented for over a century (Gibson, 1953). Such perceptual improvements, when they occur as a result of training, are called perceptual learning.

Despite its long history of research, the mechanisms of perceptual learning remain poorly understood. Instances of perceptual learning typically exhibit a number of characteristics, including specificity for stimulus parameters (e.g., spatial position and orientation of the stimulus), the simplicity of the tasks learned, and the implicit nature of the learning, that researchers have taken as evidence that perceptual learning occurs at relatively early stages of the perceptual system (Fahle & Poggio, 2002; Gilbert, 1994). Thus, many researchers interested in perceptual learning have focused on isolating changes in the neural response properties of early sensory areas following perceptual learning. While this approach has yielded results useful to understanding the neural changes underlying

certain types of non-visual perceptual learning such as vibrotactile (Recanzone, Merzenich, & Jenkins, 1992) and auditory (Recanzone, Schreiner, & Merzenich, 1993) frequency discrimination, results in visual learning tasks have been much more sparse (for reviews, see Das, 1997; Gilbert, 1994) and difficult to interpret. Furthermore, it is important to recognize that while this approach addresses questions regarding what neural changes are associated with learning, it does not answer the more central question: What is learned in perceptual learning?

One approach that has proven to be fruitful in illuminating the computational mechanisms underlying perceptual discriminations is the ideal observer framework (Geisler, 2003; Knill & Richards, 1996). This approach characterizes a given perceptual task by specifying an ideal observer, a theoretical decision-making agent described in probabilistic terms, that performs the task optimally given the available information. To determine how human observers use information in the perceptual task, researchers compare their performance with that of the ideal observer across manipulations of the task that systematically change the information available in the stimulus. This approach has been particularly successful at characterizing the ways in which observers integrate

information across different perceptual modalities (e.g., Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002; Gepshtein, Burge, Ernst, & Banks, 2005), different visual modules (e.g., Jacobs, 1999; Knill, 2003; Knill & Saunders, 2003), or both (e.g., Atkins, Fiser, & Jacobs, 2001; Hillis, Ernst, Banks, & Landy, 2002) to make perceptual judgments when multiple cues are available. Briefly, in making quotidian perceptual judgments, observers usually have access to a number of perceptual cues. An observer attempting to determine the curvature of a surface, for example, may have access to cues based on visual texture, binocular disparity, and shading, as well as to haptic cues obtained by manually exploring the surface. To make an optimal judgment based on these cues, the observer must combine the curvature estimates from these different cues. Yuille and Bülthoff (1996) demonstrated that, given certain mathematical assumptions, the optimal strategy for combining estimates $\hat{\theta}_1, \dots, \hat{\theta}_n$ from a set of class-conditionally independent cues (i.e., cues c_1, \dots, c_n that are conditionally independent given the scene parameter of interest so that $P(c_1, \dots, c_n|\theta) = \prod_i P(c_i|\theta)$) $\hat{\theta}^* = \sum_i \omega_i \hat{\theta}_i$ (where $\hat{\theta}^*$ represents the optimal estimate based on all available cues) such that the weight for each cue is inversely proportional to the variance of the distribution of the scene parameter given the cue's value (i.e., $\omega_i \propto 1/\sigma_i^2$). Researchers have found that, across a variety of perceptual tasks, human observers seem to base their perceptual judgments on just such a strategy. While most of these cue integration studies have focused on strategies used by observers in stationary environments, several (Atkins et al., 2001; Ernst, Banks, & Bülthoff, 2000; Jacobs & Fine, 1999) have investigated how observers change their cue integration strategies after receiving training in virtual environments in which a perceptual cue to a scene variable is artificially manipulated to be less informative with respect to that variable. In one of these studies, Ernst et al. (2000) manipulated either the texture- or disparity-specified slant of a visually presented surface to indicate a slant value that was uncorrelated with the haptically defined orientation of the surface. The authors found that after receiving training in this environment, subjects' perceptions of slant changed such that, in a qualitatively similar fashion to the ideal observer, they gave less weight to the slant estimate of the now less reliable visual cue.

This ideal observer framework has thus been useful in characterizing the mechanisms involved in learning to make certain types of perceptual discriminations. However, not all perceptual learning tasks fit neatly into the cue combination framework described above. Many studies of perceptual learning, for example, have focused on improvements in simple tasks involving Vernier acuity, texture discrimination, line bisection, orientation discrimination, and other *image-based* (i.e., rather than 3D scene-parameter-based) discriminations. To characterize the learning obtained in such tasks using the ideal observer cue combination framework described above, we must

first deal with several conceptual and methodological issues. The first of these issues concerns the seemingly disparate nature of 3D cue combination tasks on the one hand, and simple image-based discrimination tasks on the other. Consider for example the slant discrimination task described in the previous paragraph. In this case, the slant of the surface is defined visually by two conventional and well-understood cues to surface slant: texture foreshortening, and binocular disparity. In a texture discrimination task, however, subjects are not trying to determine the value of some surface parameter such as slant. Instead, they must determine to which of two arbitrarily defined categories a presented texture belongs. What are the cues in this task? Of course, these textures will differ along some set of image features and the subject can identify and use these features as “cues” to the texture category. But do such features function as cues in the same sense as texture foreshortening and binocular disparity? The current study was designed to address this question. We were interested in determining whether the optimal integration of cues described in cue combination studies such as that of Ernst et al. (2000) is a special property of the limited set of conventionally defined visual cues (e.g., texture compression and disparity gradient cues for slant) or whether people are likewise sensitive and capable of exploiting the relative reliabilities of arbitrarily defined cues such as those of low-level features involved in image-based discriminations.

To answer this question, we introduce an efficient modification of the classification technique that allows us to analyze over relatively fine time scales the changes to the weights an observer gives to different features. We then report the results of two experiments that exploit this technique to examine how observers use information about the reliabilities of low-level image features in performing simple perceptual discrimination tasks. Using our modified classification image technique, we investigate whether observers use information in a manner consistent with *optimal feature combination* (i.e., in a manner analogous to optimal cue combination). In both experiments, subjects viewed and classified stimuli consisting of noise-corrupted images. The stimuli used in each experiment were generated within a 20-dimensional feature space whose noise covariance structure was varied across conditions. In **Experiment 1**, subjects were trained to discriminate between two stimuli corrupted with white Gaussian feature noise, and their classifications were calculated over time. When we examined their classification images, we found that, with practice, their classification images approached that of the ideal observer. In addition, this improvement in their classification images correlated highly with their increase in performance efficiency, accounting for most of the variance in their performance. In **Experiment 2**, the variance of the corrupting noise was made anisotropic, such that some features were noisier and thus less reliable in determining the stimulus class than others. In the first half of the

experiment, half of the features were made reliable and the other half unreliable. In the second half of the experiment, this relationship was reversed so that the features which had heretofore been reliable were now unreliable and vice versa. When we examined the classification images calculated for each subject over time, we found that they modified their decision strategies in a manner consistent with optimal feature combination, giving higher weights to reliable features and lower weights to unreliable features. The results of [Experiment 1](#) suggest that subjects' learning in these texture discrimination tasks consists primarily of improvements in the optimality of their discriminant functions, while the results of [Experiment 2](#) suggest that in learning these discriminant functions, subjects are able to exploit information about the reliabilities of individual features.

Estimating classification images: A modified approach

Ahumada (1967, 2002) suggested a method for determining the template, or *classification image*, used by human observers performing a binary perceptual discrimination task. To discover this template T for an individual observer, the researcher adds random pixel noise $\epsilon^{(t)} \sim N(0, \mathbf{I})$ to the signal $\mathbf{s}^{(t)} \in \{\mathbf{s}_0, \mathbf{s}_1\}$ presented on each trial t . The researcher can then calculate the observer's classification image by simply correlating the noise added on each trial with the classification $r^{(t)} \in \{-1, 1\}$ indicated by the observer. These classification images reveal the stimulus components used by observers in making perceptual discriminations. Over the past decade, this classification image technique has proven quite useful; researchers have used this technique (or variants thereof) to determine the templates used by observers in a variety of different tasks (e.g., Abbey & Eckstein, 2002; Ahumada, 1996; Levi & Klein, 2002; Lu & Liu, 2006), to compare these observer classification images to those calculated for an ideal observer (optimal templates), and to investigate how these classification images change with learning (e.g., Beard & Ahumada, 1999; Gold, Sekuler, & Bennett, 2004). Despite these successes, the method does suffer from some shortcomings.

Chief among these is the enormous dimensionality of the stimulus space. Calculating the classification image for a stimulus represented within a 128×128 pixel space, for example, requires calculating 16,385 parameters (i.e., 128^2 regression coefficients plus a bias term). Consequently, researchers require thousands of trials to obtain a reasonable classification image for a single observer, and the correlation of the resulting images with the optimal templates is generally quite low due to the poor sampling of the stimulus space and the concomitant paucity of data points (Gold et al., 2004). Several researchers have

attempted to remedy this problem and to boost the significance of such comparisons by restricting the final analysis to select portions of the classification image (e.g., Gold et al., 2004) by averaging across regions of the image (e.g., Abbey & Eckstein, 2002; Abbey, Eckstein, & Bochud, 1999) or by using a combination of these methods (e.g., Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005). Such measures work by effectively reducing the dimensionality of the stimulus space so that instead of calculating regression coefficients for each pixel, researchers calculate a much smaller number of coefficients for various linear combinations of pixels. Essentially, these researchers add the signal corrupting noise in pixel space but perform their analyses in terms of a lower dimensional basis space.

In the current study, we simplify this process by specifying this lower dimensional basis space explicitly and *a priori*.¹ In addition to its simplicity, this approach has several advantages over traditional methods. First, by specifying the bases in advance, we can limit the added noise ϵ to the subspace spanned by these bases, ensuring that (1) the noise is white and densely sampled in this subspace, and (2) only features within the spanned subspace contribute to the observer's decisions (i.e., because all stimulus variance is contained within this subspace). Second, because we specify the bases in advance, we can select these bases in an intelligent way, representing only those features that observers are likely to find useful in making discriminations, such as those features that contain information relevant to the task (i.e., features that vary across the stimulus classes).² Finally, this approach makes it possible to manipulate the variance of the noise added to different features and thus to vary the reliabilities of these features. This allows us to investigate how observers combine information from different features using methods similar to those that have been used in studying perceptual cue combination.

Mathematically, our approach to classification images is related to Ahumada's (2002) approach as follows: let $\mathbf{g}^{(t)}$ represent the stimulus presented on trial t . Ahumada's technique generates these stimuli as

$$\mathbf{g}^{(t)} = \mathbf{s}^{(t)} + \epsilon^{(t)}, \quad (1)$$

where $\mathbf{s}^{(t)}$ and $\epsilon^{(t)}$ are defined as above. If we explicitly represent the use of pixels as bases using the matrix \mathbf{P} , whose columns consist of the n -dimensional set of standard bases, we can rewrite [Equation 1](#) in a more general form as

$$\mathbf{g}^{(t)} = \mathbf{P}(\mathbf{s}^{(t)} + \epsilon^{(t)}). \quad (2)$$

This is possible because \mathbf{P} is equivalent to the identity matrix \mathbf{I}_n . It should be clear, however, that by applying the appropriate linear transformation $T: \mathbf{P} \rightarrow \mathbf{B}$ to the stimuli

$\mathbf{s}^{(t)}$, we can exchange \mathbf{P} for an arbitrary basis set \mathbf{B} to generate stimulus images in the space spanned by \mathbf{B} . This is represented by our generative model

$$\mathbf{g}^{(t)} = \mathbf{k} + \mathbf{B}(\boldsymbol{\mu}^{(t)} + \boldsymbol{\eta}^{(t)}), \quad (3)$$

where $\boldsymbol{\mu}^{(t)} \in \{\boldsymbol{\mu}_A, \boldsymbol{\mu}_B\}$ represents a prototype stimulus expressed in terms of the basis set, and $\boldsymbol{\eta}^{(t)} \sim N(0, \mathbf{I})$ represents Gaussian noise added in the basis space. (Note that when $\mathbf{B} = \mathbf{P}$, Equation 3 is equivalent to Equation 2, with $\boldsymbol{\mu}^{(t)} = \mathbf{s}^{(t)}$, $\mathbf{k} = 0$, and $\boldsymbol{\eta}^{(t)}$ and $\boldsymbol{\epsilon}^{(t)}$ distributed identically.) The only new term is the constant vector \mathbf{k} , which is important here because it provides additional flexibility in choosing the bases that make up \mathbf{B} .³ In particular, this constant term allows us to represent constant (noiseless) features in pixel space that do not exist in the space spanned by \mathbf{B} . Figures 1 and 2 illustrate this generative model for a pair of example stimuli. Here the task requires classifying a presented stimulus as an instance of stimulus A (square) or stimulus B (circle). All of the information relevant to this discrimination lies in the difference image (the rightmost image in Figure 1). The image shown to the left of this difference image (third from left) represents the part of the stimulus that remains constant across stimulus classes. Representing this part of the stimulus as \mathbf{k} allows us to focus on selecting bases \mathbf{B} that can adequately represent the difference image. Figure 2 shows example stimuli \mathbf{g} generated for this task using the models described in Equation 2 (top of Figure 2) and Equation 3 (bottom of Figure 2).

The method developed by Ahumada for calculating classification images is—despite its successful use by many researchers—somewhat inaccurate and can potentially be quite inefficient. Ahumada’s method is based on reverse correlation, a technique for determining the linear response characteristics of a signal processing system. In reverse correlation, a researcher feeds Gaussian white noise into a system, records the system’s output, and then characterizes the system’s linear response by correlating the input and output signals. Unfortunately, however, psychophysical experiments that use the classification image technique rarely present pure noise to observers in practice because this tends to result in unreliable performance (for a counterexample, see Neri & Heeger,

2002). Instead, they typically corrupt one of two signals (where one of the signals may be the null signal) with noise and have the observer determine which of the two signals was presented. As a result, the observer is actually exposed to signals from two distributions with different means rather than just one. Ahumada’s method for dealing with this problem is to subtract the means from these two distributions (Ahumada, 2002) and thereafter treat them as a common distribution. At best, ignoring the signal and considering only the noise makes for an inefficient estimate of the observer’s decision template since it ignores available information. At its worst, ignoring the signal can lead to some rather strange results (consider, for example, that subjects who perform at 50% correct and at 100% correct are indistinguishable using this method).

Since one of our goals in this study was to develop a more efficient means of estimating classification images, we calculated the maximum likelihood estimate for these images using the full stimuli (signal + noise) under a Bernoulli response likelihood model. Here, we show that the classification image for the ideal observer (the optimal template) can be expressed as the result of a logistic regression. We assume that the ideal observer knows the prior distributions $P(C_i)$ and likelihood functions $P(\mathbf{x}|C_i)$ for both stimulus classes C_i , $i \in \{A, B\}$. Using Bayes’ rule, the probability that an image \mathbf{x} belongs to class A is

$$\begin{aligned} P(C_A|\mathbf{x}) &= \frac{P(\mathbf{x}|C_A)P(C_A)}{P(\mathbf{x})} \\ &= \frac{P(\mathbf{x}|C_A)P(C_A)}{P(\mathbf{x}|C_A)P(C_A) + P(\mathbf{x}|C_B)P(C_B)}, \end{aligned} \quad (4)$$

With some simple algebra, we can convert this expression into a logistic function of \mathbf{x} .

$$P(C_A|\mathbf{x}) = \frac{1}{1 + e^{-f(\mathbf{x})}}, \quad (5)$$

where

$$f(\mathbf{x}) = \log \left[\frac{P(\mathbf{x}|C_A)P(C_A)}{P(\mathbf{x}|C_B)P(C_B)} \right]. \quad (6)$$

To express the classification image as the result of a logistic regression however, we must also demonstrate

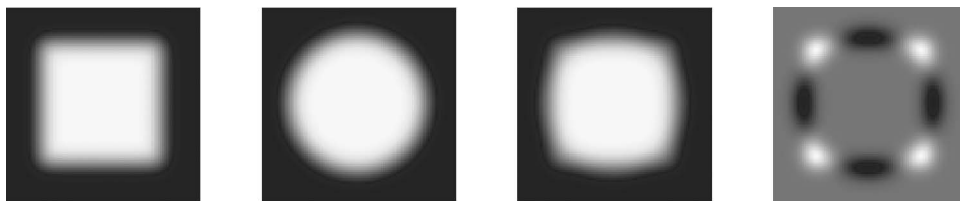


Figure 1. An illustrative stimulus set consisting of “fuzzy” square and circle prototypes. From left to right: the square ($\mathbf{k} + \mathbf{B}\boldsymbol{\mu}_A$); the circle ($\mathbf{k} + \mathbf{B}\boldsymbol{\mu}_B$); the constant image (\mathbf{k}), which represents the parts of the image that are invariant across stimuli, and the square–circle difference image ($\mathbf{B}[\boldsymbol{\mu}_A - \boldsymbol{\mu}_B]$).

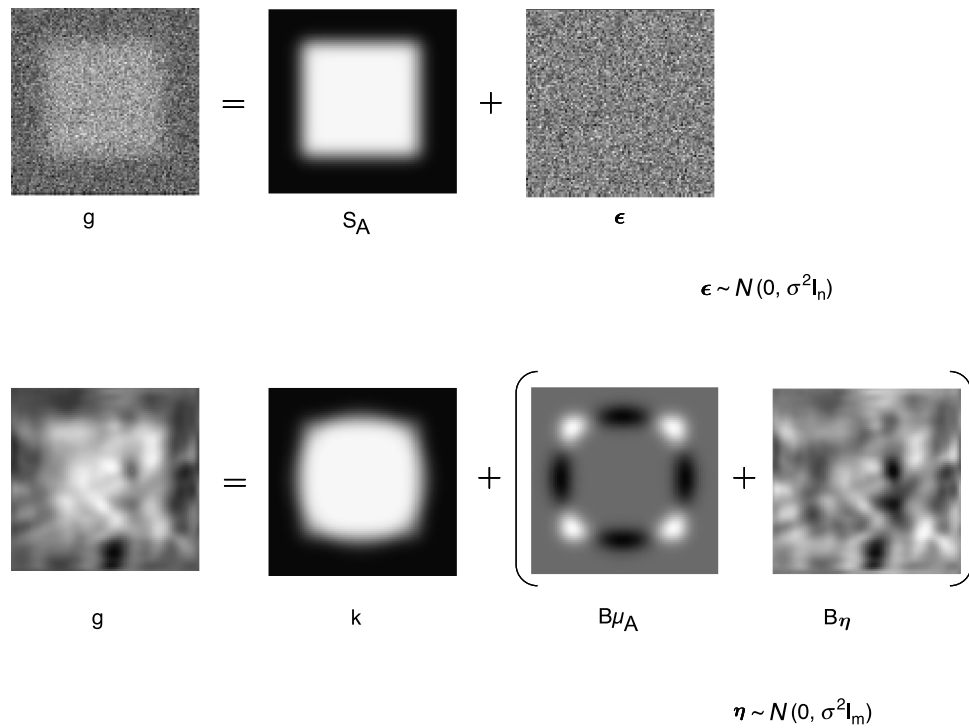


Figure 2. Illustrations of the methods described in Equation 2 (top) and Equation 3 (bottom) for generating noise-corrupted versions of the “fuzzy square” prototype (stimulus A) introduced in Figure 1.

that $f(\mathbf{x})$ in Equation 6 is linear in \mathbf{x} . The stimuli presented on each trial are drawn from a multivariate Gaussian representing one of the two signal categories. Therefore, we can express the likelihood terms in Equation 6 as

$$P(\mathbf{x}|C_i) = (2\pi)^{-\frac{m}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma^{-1} (\mathbf{x}-\mu_i)}, \quad (7)$$

where μ_i , $i \in A, B$ is the mean (prototype) for class i , Σ is the common covariance matrix for both classes, and m is the dimensionality of the stimulus space. Plugging these likelihoods into Equation 6 yields

$$f(\mathbf{x}) = \frac{1}{2} \left[(\mathbf{x}-\mu_B)^T \Sigma^{-1} (\mathbf{x}-\mu_B) - (\mathbf{x}-\mu_A)^T \Sigma^{-1} (\mathbf{x}-\mu_A) \right] + \log \left[\frac{P(C_A)}{P(C_B)} \right]. \quad (8)$$

Finally, by expanding the quadratic terms and simplifying, we demonstrate that $f(\mathbf{x})$ is indeed linear in \mathbf{x} :

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad (9)$$

with

$$\mathbf{w} = \Sigma^{-1} (\mu_A - \mu_B), \quad (10)$$

and

$$b = \frac{1}{2} (\mu_B + \mu_A)^T \Sigma^{-1} (\mu_B - \mu_A) + \log \left[\frac{P(C_A)}{P(C_B)} \right]. \quad (11)$$

Equation 10 shows that in the case of white Gaussian noise (i.e., when $\Sigma = \sigma^2 \mathbf{I}$) the optimal template is proportional to the difference between the signal category prototypes. Note also the similarity of Equation 10 to the result $w_i \propto 1/\sigma_i^2$ from optimal cue combination. We exploit this relationship in the design of Experiment 2.

Experiment 1

In Experiment 1, we calculated response classification images for observers learning to perform an image-based perceptual discrimination task. We expected that our subjects’ performances would improve over time and, based on the results of Gold et al. (2004), that improvements in a subject’s discrimination performance would be accompanied by an increased fit between the observer’s classification image and the ideal template. In addition, we expected that

constructing our stimuli from a small set of bases would allow us to calculate robust classification images using a significantly smaller number of trials than are required by the traditional approach of using image pixels as bases.

Methods

Subjects

Subjects were four students at the University of Rochester with normal or corrected-to-normal vision. All subjects were naive to the purposes of the study.

Stimuli

The stimuli were 256×256 pixel ($8^\circ \times 8^\circ$) gray scale images presented on a gray background whose luminance of 16.5 cd/m^2 matched the mean luminance of the images. All of the stimuli were constructed as linear combinations of the set of basis “features” illustrated in [Figure 3](#).

The set of 20 basis features was constructed in the following manner. We created 50 32×32 pixel images of white gaussian noise which were band-pass filtered to contain frequencies in the range of 1–3 cycles per image. The resulting images were then iteratively adjusted using gradient descent to yield a set of orthogonal, zero-mean images that maximized smoothness (i.e., minimized the sum of the Laplacian) across each image. The images were added to the basis set one by one so that each basis provided an additional orthogonality constraint on the subsequent bases. In other words, at iteration i , image i was modified via gradient descent to be maximally smooth and to be orthogonal to images 1 through $(i - 1)$. These orthogonality constraints interacted with the smoothness constraint to produce images that were localized in spatial frequency content such that the first bases produced by our method contained low frequencies and subsequently added bases contained increasingly higher frequencies. We randomly selected twenty of the 50 images to form the basis set that we used to construct our stimuli.

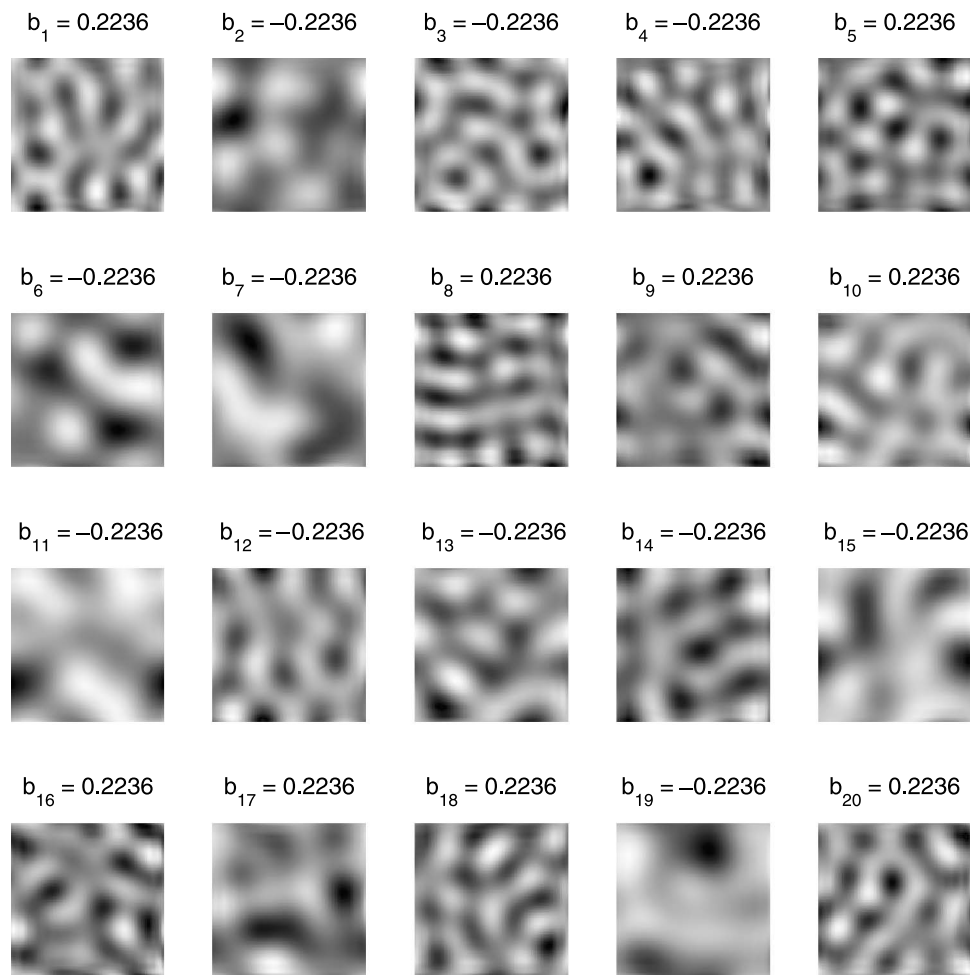


Figure 3. The 20 basis features used to construct the stimuli in [Experiments 1](#) and [2](#). Each of these images constitutes a column of the matrix \mathbf{B} in [Equation 3](#). Mixing coefficients μ_{A_i} for the vector μ_A representing Prototype A (see [Figure 4](#)) are indicated above each of the bases ($\mu_{B_i} = -\mu_{A_i}$). White Gaussian noise (in the subspace spanned by \mathbf{B}) is generated by independently sampling the noise coefficients η_i from a common Gaussian distribution.

Finally, we wanted to make sure that the bases were equally salient. The human visual system is known to exhibit varying sensitivity to different stimuli depending on their spatial frequency content. This differential sensitivity across spatial frequencies is often characterized through the contrast sensitivity function (CSF), which describes the amount of contrast required at different spatial frequencies to obtain a fixed sensitivity level. Thus, as a final step, we normalized the twenty basis features for saliency by setting the standard deviation of luminance distributions in each of the basis images to 1, then multiplied each image by the reciprocal of the contrast sensitivity function value for its peak spatial frequency.⁴

A set of two prototypes was constructed from this basis set as follows. First, a 20-dimensional vector was formed by randomly setting each of its elements to either 1.0 or -1.0 . The result was an image centered within one of the orthants of the space spanned by the basis features. This vector represented prototype A. The vector representing the second prototype, prototype B, was simply the negative of the vector representing prototype A ($\mu_B = -\mu_A$). To obtain images of the prototypes, these vectors were multiplied by the matrix representing the 20 basis features and a constant image was added, consisting of the mean luminance plus an arbitrary image constructed in the null space of the basis set (the addition of this arbitrary image prevented the prototypes from appearing simply as contrast-reversed versions of the same image). Finally, the prototypes were upsampled to yield 256×256 pixel images. We created only one set of prototypes and all subjects saw the same set (Figure 4).

Test stimuli were created according to the generative model described in Equation 3. On each trial, one of the two prototypes (A or B) was selected at random and combined with a noise mask $\eta^{(t)}$. The noise masks, like the prototypes, were generated as a linear combination of the basis features. However, for the noise masks, the linear coefficients were sampled from a multivariate Gaussian distribution $\eta^{(t)} \sim N(0, \sigma^2 \mathbf{I})$. Values that deviated more than 2σ from the mean were resampled. The RMS contrast of the signal and the noise mask were held constant at 5.0% and 7.5%, respectively.

Procedure

Each trial began with the presentation of a fixation square, which appeared for 300 ms. This was followed by

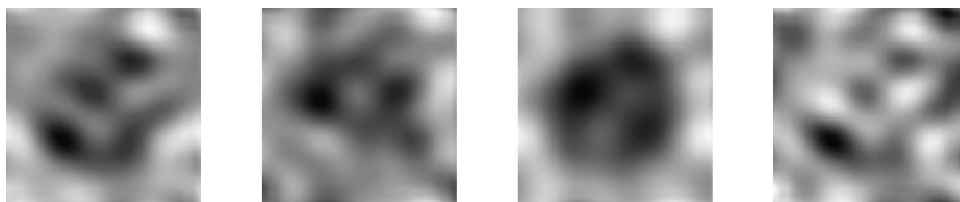


Figure 4. The prototypes used in Experiments 1 and 2 presented in the same format as the example stimuli in Figure 1. From left to right: prototype A ($\mathbf{k} + \mathbf{B}\mu_A$), prototype B ($\mathbf{k} + \mathbf{B}\mu_B$), the constant image (\mathbf{k}), and the difference image ($\mathbf{B}[\mu_A - \mu_B] = 2\mathbf{B}\mu_A$).

a test stimulus, which was also presented for 300 ms. Both the fixation square and the test stimulus were centered on the screen. One hundred fifty milliseconds after the test stimulus had disappeared, the two prototypes were faded in, laterally displaced 8° (256 pixels) from the center of the screen. Subjects were instructed to decide which of the two prototypes had appeared in the test stimulus and responded by pressing the key corresponding to the selected prototype. Subjects received immediate auditory feedback after every trial indicating the correctness of their response. In addition, after every 15 trials, a printed message appeared on the screen indicating their (percent correct) performance on the previous 15 trials. Each subject performed 12 sessions of 300 trials each over 3 days, and the subject's response, the signal identity, and the noise mask were saved on each trial to allow calculation of the subject's classification image.

Results

Figures 5 and 6 and Table 1 summarize the results of this experiment. We wanted to determine the following:

1. Can subjects learn to discriminate texture stimuli generated in our basis space?
2. How well do improvements in discrimination performance correlate with the optimality of an observer's classification image?
3. How efficient is our method? That is, how many trials are required to estimate a subject's classification image?

To determine whether our observers learned in this task, we correlated their sensitivity d' in each session with the total number of trials completed at the end of that session.

The results of this correlation across the 12 sessions are shown in Table 1. Three of the four subjects showed significant improvement between the first and second halves of training, indicating that subjects could indeed learn to discriminate stimuli in our basis space. We calculated classification images for each session using logistic regression (see Equations 4–11).

Figure 5 shows classification images obtained over the first and last quarter of trials for each of the three subjects who showed learning. There are clear changes to the

Subject	$r(df)$	p
WHS	$r(10) = 0.7657$	<0.005
RAW	$r(10) = 0.8518$	<0.001
BVR	$r(10) = 0.8126$	<0.005
SKL	$r(10) = 0.3745$	>0.05

Table 1. Correlation between sensitivity and trial number for individual subjects.

images as a result of learning. To quantify these changes, we calculated the normalized-cross-correlation ($\mathbf{w}_{\text{obs}}^T \mathbf{w}_{\text{ideal}} / (||\mathbf{w}_{\text{obs}}|| ||\mathbf{w}_{\text{ideal}}||)$) between the subject's classification image \mathbf{w}_{obs} and that of the ideal observer $\mathbf{w}_{\text{ideal}}$ across time. Normalized-cross-correlation is often used to represent the degree of “fit” between two templates (e.g., Gold et al., 2004; Murray, 2002). The “fit” in this case is indicative of the optimality of the template used by a particular subject, and we thus refer to the square of the normalized cross-correlation as the subject's *template efficiency* (Figure 6, dashed curve). We also calculated subjects' discrimination efficiencies $[d'_{\text{obs}}/d'_{\text{ideal}}]^2$ (Geisler, 2003) for each session to compare the performances of subjects to that of the ideal observer. Finally, we correlated each subject's discrimina-

tion and template efficiencies across sessions to measure how improvements in discrimination performance correlate with improvements in the optimality of the subject's classification image. The resulting correlation coefficients and significance statistics appear at the top of the plots in Figure 6. The correlations are quite strong, indicating that increases in subjects' discrimination efficiencies are well explained by the observed improvement in their templates. This finding corroborates a qualitatively similar finding by Gold et al. (2004).

Overall, the results of Experiment 1 demonstrate that our method for obtaining and calculating classification images represents a successful improvement over existing methods for studying perceptual learning. Our use of arbitrary basis features did not preclude learning. Limiting the number of features, however, allowed us to calculate subjects' classification images over short time scales (<300 trials) and thus to track changes in subjects' templates throughout the course of learning. Additionally, the results suggest that most of the variance in subjects' discrimination performances (i.e., 66% to 81%)⁵ can be accounted for by improvements in their classification images, so that changes in subjects' discrimination strategies over time can largely be characterized by calculating their classification images. Together, these

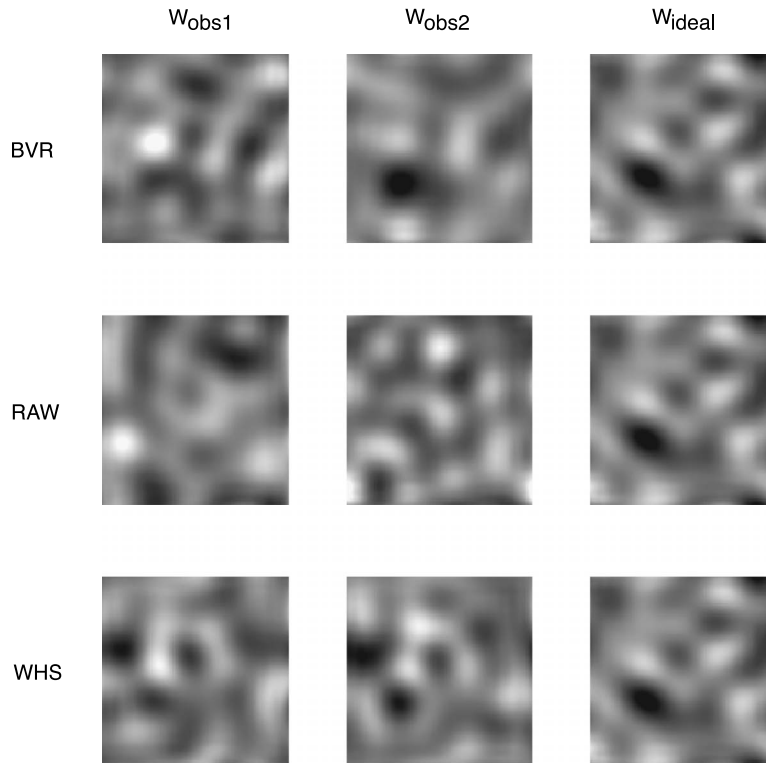


Figure 5. Classification images for each of the three subjects who showed learning in Experiment 1. The first column \mathbf{w}_{obs1} displays the subjects' classification images calculated over the first three sessions; the second column \mathbf{w}_{obs2} displays the classification images calculated over their final three sessions; and the third column $\mathbf{w}_{\text{ideal}}$ displays the optimal template.

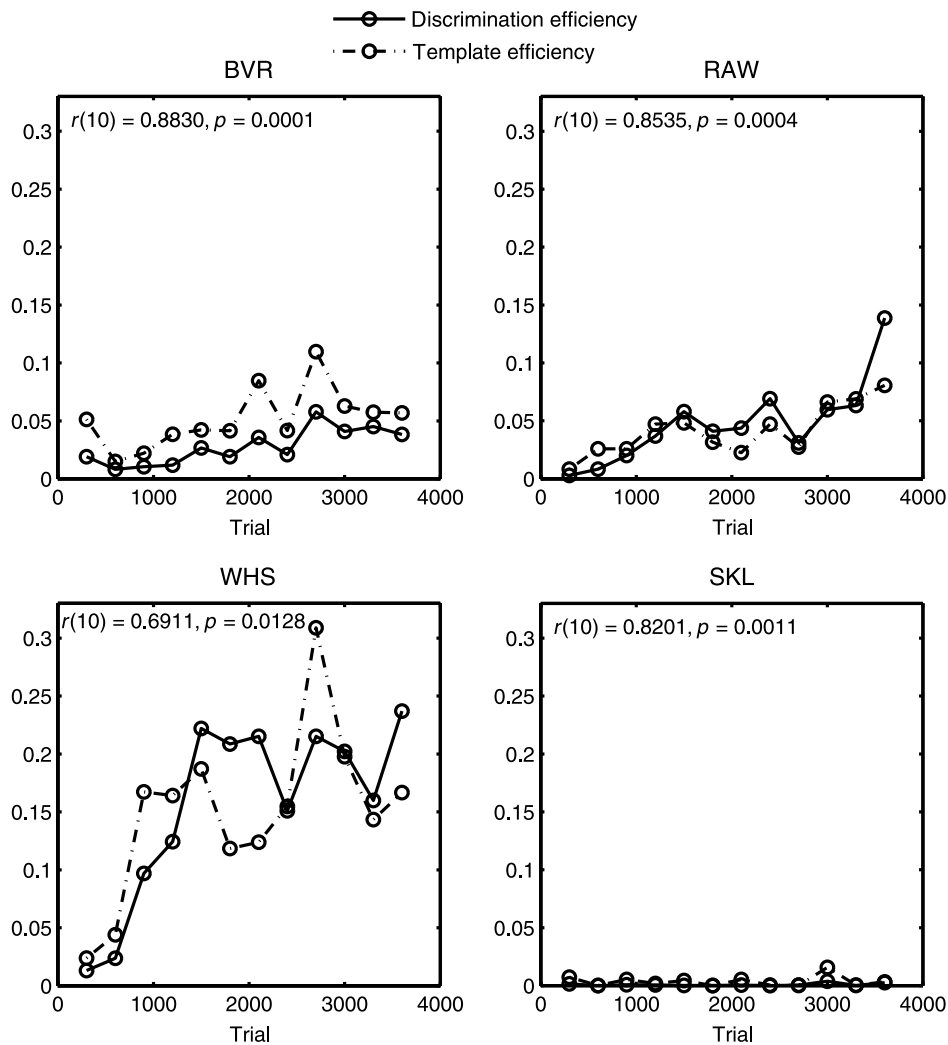


Figure 6. Individual results for all 4 subjects who participated in [Experiment 1](#). The horizontal axis of each plot indicates the trial number, while the vertical axis represents both the subject's discrimination efficiency (solid curve) and template efficiency (dashed curve). The correlation coefficient for the fit between these two measures and the p -value representing the significance of this correlation is indicated at the top of each subject's plot.

characteristics indicate that our method is suitable for determining how observers change their discrimination strategies as a perceptual task is modified.

Experiment 2

[Experiment 2](#) was designed to determine whether observers modify their templates in a manner consistent with *optimal feature combination* (i.e., in a manner analogous to optimal cue combination). We investigated this question by manipulating the reliabilities of different features with respect to discrimination judgments like those made by subjects in [Experiment 1](#). Changes made to the relative reliabilities of different features result in

corresponding changes to the optimal decision template. By calculating the classification images used by subjects across such manipulations, we can determine whether observers are sensitive to the reliabilities of individual features and modify their templates accordingly. The idea, illustrated in [Figures 7B](#) and [7C](#), is to change the optimal template across two phases of the experiment by modifying only the variance structure of the noise. If observers use information about feature variance in performing discrimination tasks, then we should observe a change in their classification images between the first and the second phases of the experiment. After the transition, observers' templates should move away from that predicted by the optimal template for the first set of reliable versus unreliable features, and toward that predicted by the optimal template for the second set. We expected that subjects would take feature reliabilities into account when

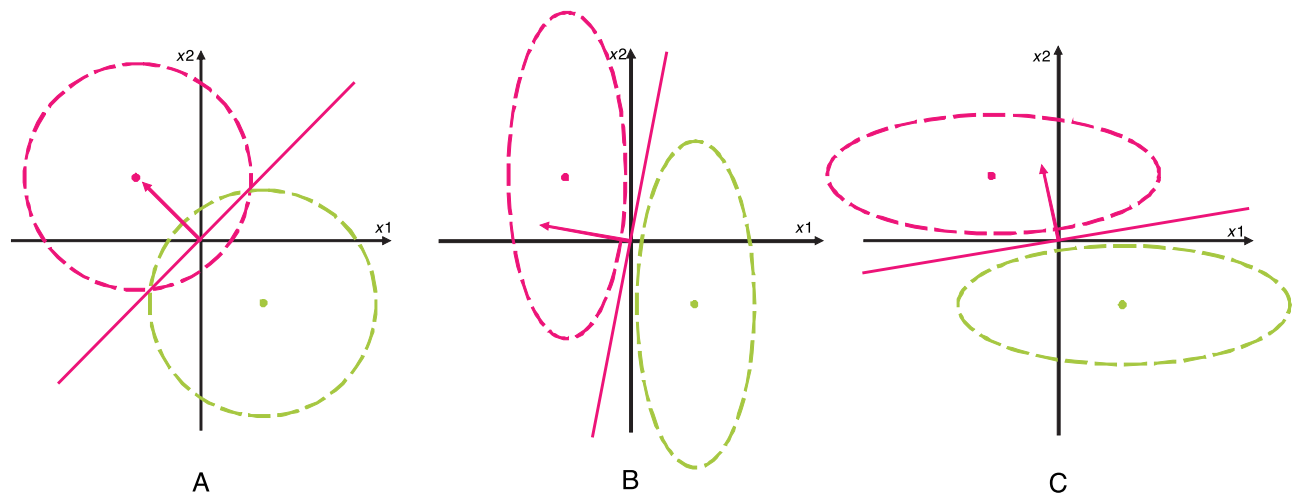


Figure 7. A schematic illustration of the effect of variance structure on the optimal template (red arrows) for a two-dimensional stimulus space. Dashed lines represent contours of equal likelihood ($P(x_1, x_2|C_i) = k$) for category A (red) and category B (green). The solid red lines and arrows represent the optimal decision surface and its normal vector (i.e., the template for category A), respectively. (Left) Two prototypes embedded in isotropic noise ($\Sigma = I_2$). (Center) The variance along dimension x_2 is greater than that in x_1 . (Right) The variance along x_1 is greater than that in x_2 .

making discriminations, resulting in classification images that give greater weight to reliable features and lower weight to unreliable features.

Methods

Subjects

Subjects were four students at the University of Rochester with normal or corrected-to-normal vision. All subjects were naive to the purposes of the study.

Stimuli and procedure

The task for observers in this experiment was identical to the task described in [Experiment 1](#). Observers classified a briefly presented stimulus as an instance of either stimulus A or stimulus B. Prototypes A and B were also identical to those used in [Experiment 1](#), and the stimuli for each trial were constructed according to the generative model described in [Equation 3](#), except that the noise covariance matrix Σ was not the identity matrix. Observers performed 24 sessions of 300 trials each over 6 days.

The procedure for [Experiment 2](#) differed from that of [Experiment 1](#) in that [Experiment 2](#) consisted of two phases, each comprising 12 sessions. Before training, 10 of the 20 basis features were selected at random to be “unreliable” features, so that each subject had a unique set of reliable and unreliable features. We controlled the reliability of an individual feature b_i by manipulating its variance σ_i^2 in the noise covariance matrix Σ . [Equation 10](#) establishes the relationship between the noise covariance and the optimal template. Exploiting the facts that Σ is a

diagonal matrix and that $\mu_B = -\mu_A$, we can express the individual elements of \mathbf{w} as

$$w_i = \frac{2\mu_{A_i}}{\sigma_i^2}, \quad (12)$$

where σ_i^2 represents the i th diagonal element of Σ . Note that this is similar to the result obtained for optimal weighting of independent cues in the literature on cue combination (e.g., Landy, Maloney, Johnston, & Young, 1995; Yuille & Bülthoff, 1996). The difference here is that instead of simply weighting each feature in proportion to its reliability (i.e., inverse variance), there is an added dependency on the class means, such that observers must weight each feature in proportion to its mean-difference-weighted reliability. In the current study, we removed this dependency by choosing the elements in μ_A such that their magnitudes are all equal (i.e., $|\mu_{A_i}| = |\mu_{A_j}| \forall i, j \leq m$) so that the weights composing the optimal template are indeed inversely proportional to the variances of their associated features.⁶ [Figure 7](#) illustrates this dependency for a simple stimulus space consisting of two feature dimensions x_1 and x_2 .

In the first half of training (sessions 1–12), the variance of the noise added to the unreliable features was greater than the variance of the noise added to the reliable features (i.e., $\sigma_{\text{unreliable}} = 5$ while $\sigma_{\text{reliable}} = 1$). In the second half of the experiment, the roles of these two sets of features were swapped such that the reliable features were made unreliable and the unreliable features were made reliable. Importantly, the set of reliable and unreliable features

Subject	$t(df)$	p
DLG	$t(5) = -5.7661$	<0.005
JDG	$t(5) = -3.3911$	<0.05
MSB	$t(5) = -13.3369$	<0.0001
MKW	$t(5) = -27.4861$	<0.00001

Table 2. Significance statistics for results displayed in Figure 10.

were chosen randomly for each subject so that the pair of covariance matrices for the first (Σ_1) and second (Σ_2) halves of the experiment were unique to a subject.

Results

We wanted to determine whether subjects adjusted their discrimination strategies in a manner consistent with optimal feature combination when the variance of individual features was modified. As in Experiment 1, we calculated classification images for each of our subjects and quantified the fit between these images and the templates used by an ideal observer using normalized-

cross-correlation. In contrast to Experiment 1, however, each subject made discriminations under two different generative models using covariance matrices Σ_1 and Σ_2 , respectively. Thus, we defined two optimal templates for each subject; one for the generative model used in sessions 1–12 (w_{ideal1} , appropriate for Σ_1) and one for the generative models used in sessions 13–24 (w_{ideal2} , appropriate for Σ_2). Figure 8 plots the normalized-cross-correlation between the calculated classification image w_{obs} and the templates w_{ideal1} (solid lines) and w_{ideal2} (dashed lines) for each of the four subjects as a function of the number of trials. Figure 9 displays the visible change in the classification images used by subjects between the first and second halves of the experiment.

These plots demonstrate that subjects modified their decision templates in accordance with our predictions, employing templates that fit more closely with w_1 when the noise covariance structure was defined by Σ_1 and modifying their templates to more closely match w_2 during the second half of the experiment, when the covariance structure was defined by Σ_2 . To quantify these results, we compared the average difference between the template fits $wfit_2 - wfit_1$ (where $wfit_i$ represents the normalized-cross-correlation between template w_i and a

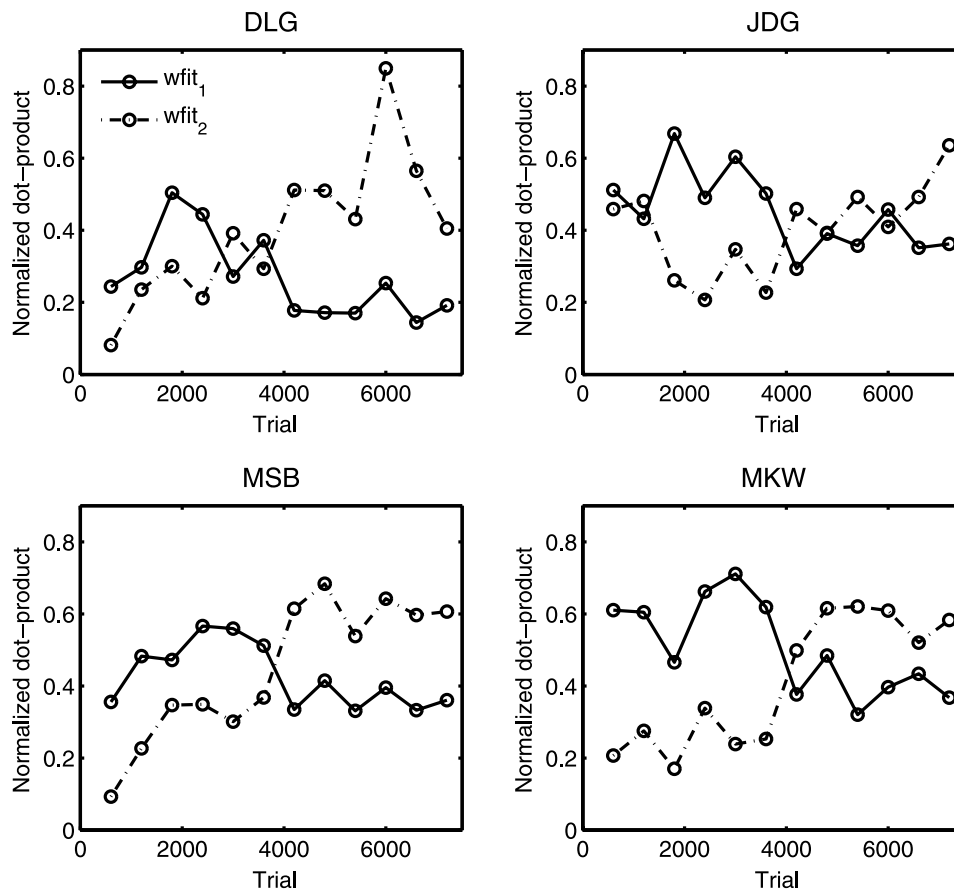


Figure 8. Normalized cross-correlation for each of the four subjects in Experiment 2. The plots depict the fits between each subject's classification image (w_{obs}) and the optimal templates for the covariance structure of the noise used in the first (solid lines) and second (dashed lines) halves of the experiment. The change in covariance structure occurred at trial 3601.

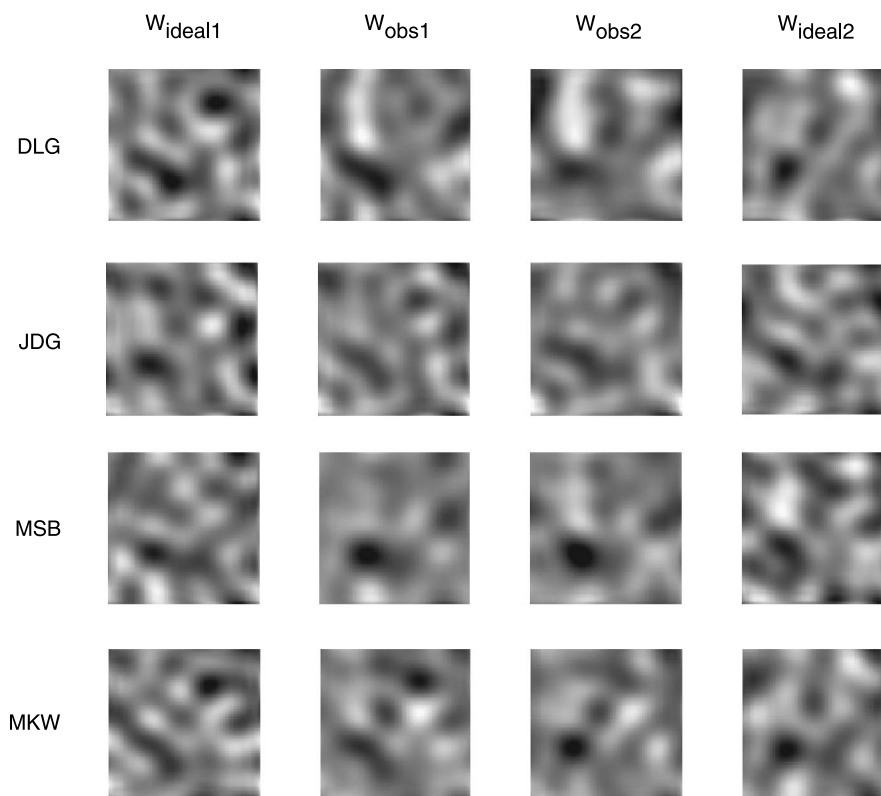


Figure 9. Classification images for each of the four subjects in Experiment 2. The first column displays the optimal template w_{ideal1} calculated for the feature covariance Σ_1 used in the first half of the experiment; the second column w_{obs1} displays the subjects' classification images calculated over the first 12 sessions; the third column w_{obs2} displays the classification images calculated over their final 12 sessions; and the final column displays the optimal template w_{ideal2} calculated for the feature covariance Σ_2 used in the second half of the experiment.

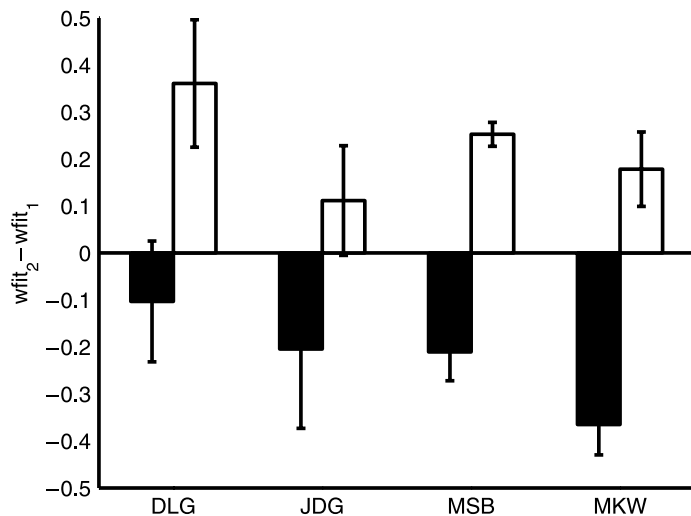


Figure 10. The differences between the template fits ($wfit_2 - wfit_1$) plotted in Figure 8 averaged over the first (open bars) and second (closed bars) half of trials in Experiment 2.

subject's classification image) across the first and second halves of the experiment using a t -test. These differences are plotted in Figure 10 and the corresponding significance statistics displayed in Table 2.

In summary, using the methods introduced in Experiment 1 for obtaining and calculating classification images, Experiment 2 examined whether human observers exploit information about the reliabilities of individual features when performing an image-based perceptual discrimination task. We manipulated the reliabilities of our features by changing the covariance structure over time. Our results show that subjects change their classification images to track changes in the optimal template, suggesting that they indeed use information about the reliabilities of individual features, giving greater weight to more reliable features in a manner analogous to optimal cue combination.

Discussion

Researchers have repeatedly demonstrated that, with practice, observers can learn to significantly improve their performance in many perceptual discrimination tasks. The

nature of this learning, however, is not well understood. The two experiments described in this paper contribute to our understanding of perceptual learning by studying how observers improve their use of stimulus information as a result of practice with a discrimination task.

First, we introduced a modification of the classification image technique that through its improved efficiency allows us to track the changes to observers' templates as the result either of learning or of experimental manipulations. We investigated whether observers use information in a manner consistent with *optimal feature combination* (i.e., in a manner analogous to optimal cue combination). In both Experiments, subjects viewed and classified stimuli consisting of noise-corrupted images. The stimuli used in each experiment were generated within a 20-dimensional feature space whose noise covariance structure varied across conditions. In [Experiment 1](#), subjects were trained to discriminate between two stimuli corrupted with white Gaussian feature noise and their classifications were calculated over time. Examination of their classification images reveals that, with practice, their decision templates approached that of the ideal observer. Moreover, this improvement in their classification images correlated highly with their increase in performance efficiency, accounting for between 65% and 80% of the variance in their performance. Consistent with the findings of Gold et al. (2004), these results suggest that the learning demonstrated in these perceptual discrimination tasks consists primarily of observers improving their discriminant functions to more closely match the optimal discriminant function.

But what does it mean to say that improvements in perceptual discrimination tasks result primarily from the learning of optimal discriminant functions? Discriminant functions encode information about several distinct aspects of the stimuli to be discriminated. The first of these is the prior probability over stimulus categories. If one type of signal is more likely than another, then an optimal observer judging the category membership of an ambiguous stimulus should assign a higher probability to the more likely category. The second of these is the mean signals in each category—the category prototypes. The importance of this aspect of the stimuli for the discriminant function is obvious. Deciding which of two noise-masked signals was presented is quite difficult if the observer cannot identify the signals in the absence of a noise mask. In white noise, the optimal discriminant surface between two signal categories is perpendicular to the vector describing the difference between the category prototypes. Most studies using classification image techniques, by using white noise masks and flat category priors exclusively, have primarily examined this aspect of perceptual discriminant functions—asking how well observers represent signal prototypes in making perceptual discriminations. The third aspect of the task encoded in discriminant functions is the structure of the variance or noise in the features that define the stimuli. As demonstrated in [Equation 10](#) and in [Figure 7](#), changes made to the feature covariances can dramatically alter the optimal

template for a discrimination task. No previous work with classification images had (to our knowledge) explored how observers use information about noise structure. Thus, in [Experiment 2](#), we applied a framework previously used in cue integration studies to determine how observers use information about class-conditional variance across features in a perceptual discrimination task. We were particularly interested in determining whether observers can integrate optimally across noisy features in a manner consistent with optimal cue combination. Thus, emulating a procedure used in many cue integration experiments, we manipulated the reliabilities of different features by increasing the variance in a subset of the features to make these features unreliable. As described above, this manipulation altered the optimal template for the resulting discrimination tasks. In both [Experiments 1](#) and [2](#), subjects' classification images, with practice, approached the optimal template, demonstrating that human observers are sensitive to the variances of individual features—even when these features are chosen arbitrarily—and that they use information about these variances in making perceptual judgments. In addition, that subjects in [Experiment 2](#) changed their templates in response to changes in the reliabilities of features, giving greater weight to reliable features and less weight to unreliable features, suggests that observers use this information in a manner consistent with optimal cue combination.

In summary, our results suggest that learning in image-based perceptual discrimination tasks consists primarily of changes that drive the discriminant function used by human observers nearer to that used by the ideal observer. Moreover, in learning these discriminant functions, observers seem to be sensitive to the individual reliabilities of arbitrary features, suggesting that optimal cue integration in vision is not restricted to the combination of estimates from a set of canonical visual modules (e.g., texture and disparity-based estimators for slant) in making surface-based discriminations but is instead a more general property of visual perception that generalizes to simple image-based discrimination tasks. Although the current study only investigated feature integration in a single texture discrimination task, we believe that this task is representative of many other simple discrimination tasks. However, future research is needed determine whether our results generalize to similar training in other tasks (e.g., Vernier discrimination, motion direction discrimination, orientation discrimination).

Finally, note that the current paper uses a *normative* approach to modeling what observers learn through practice with a perceptual discrimination task. This approach focuses on the structure of the task that an observer must solve, on the relevant information available to the observer, and on the fundamental limits that these factors place on the observer's performance. In contrast to process-level models of perceptual learning (e.g., Bejjanki, Ma, Beck, & Pouget, 2007; Lu & Doshier, 1999; Otto, Herzog, Fahle, & Zhaoping, 2006; Petrov, Doshier, &

Lu, 2005; Teich & Qian, 2003; Zhaoping, Herzog, & Dayan, 2003) the normative approach used here is largely agnostic with respect to either physiological or algorithmic implementation details (Marr, 1982). Our results demonstrate that people can learn to use information about the covariance structure of a set of arbitrary low-level visual features. We leave the question of how this learning is implemented in the brain as a problem for future work.

Acknowledgments

We thank the reviewers for helpful comments on an earlier version of the manuscript. This work was supported by NIH research grant R01-EY13149 and by AFSOR grant FA9550-06-1-0492.

Commercial relationships: none.

Corresponding author: Robert A. Jacobs.

Email: robbie@bcs.rochester.edu.

Address: 416 Meliora Hall, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627-0268, USA.

Footnotes

¹Several researchers (e.g., Olman & Kersten, 2004; Li, Levi, & Klein, 2004) have previously introduced lower-dimensional methods for calculating classification images (or classification objects). Note however that the approaches used in these papers differ from the approach used in the current paper in that they obtain this reduction in dimensionality by assuming that observers have direct access to geometric scene configurations rather than to the photometric input (e.g., pixel intensities) that subjects actually observe. In Li et al. (2004), the authors implicitly assume that observers have direct access to an array whose entries represent the positions of the elements making up a Vernier stimulus and that they make decisions based on this vector of positions rather than on the pattern of luminances within the image. Similarly, Olman and Kersten (2004) assume that observers have direct access to variables describing the geometry of the scene (e.g., foot spread, tail length, tail angle, neck length). In these two studies, the stimuli are defined directly in terms of scene variables—though subjects in fact observe these variables through images—and the resulting classification images are linear in the geometrical object space, but not in image space. These approaches may be more useful than image-based approaches for investigating how observers make discriminations in tasks involving representations of three-dimensional scenes (as in Olman & Kersten, 2004) when researchers have an adequate understanding of the internal representations used by observers.

²Simoncelli, Paninski, Pillow, and Schwartz (2004) provide an extended discussion regarding the importance of stimulus selection in the white noise characterization of a signal processing system. Though they are concerned in particular with characterizing the response properties of neurons, their points apply equally well to the challenges involved in characterizing the responses of human observers in a binary discrimination task. Olman and Kersten (2004) provide a related discussion that proposes extending noise characterization techniques to deal with more abstract (i.e., non-photometric) stimulus representations.

³The constant \mathbf{k} is used to represent any constant component of the image. In fact, because luminance values cannot be negative, traditional approaches to classification images implicitly include a \mathbf{k} in the form of a mean luminance image (e.g., a vector of identical positive pixel luminance values).

⁴Contrast sensitivity functions were not measured directly for each subject. Instead, for the sake of expediency, we used the model of human contrast sensitivity proposed by Mannos and Sakrison (1974), which describes the sensitivity of a human observer, generically, as $A(f) = 2.6(0.0192 + 0.114f)e^{-(0.114f)^{1.1}}$.

⁵These estimates of explained variance are obtained using the correlation between the normalized cross correlations ($\mathbf{w}_{\text{obs}}^T \mathbf{w}_{\text{ideal}} / \|\mathbf{w}_{\text{obs}}\| \|\mathbf{w}_{\text{ideal}}\|$) and the sensitivity ratio $[d'_{\text{obs}}/d'_{\text{ideal}}]^2$. Unlike in Figure 6, these values were not squared. Squaring the sensitivity measure is necessary for an information-theoretic interpretation of efficiency, but removes information about some of the correlation between observers' template fits and sensitivities (e.g., classification images that point in the wrong direction yield sensitivities below zero). The r^2 values resulting from this correlation are: 0.80 (BVR), 0.76 (RAW), 0.66 (WHS), and 0.81 (SKL).

⁶In general, if the stimuli are not chosen arbitrarily, $|\mu_A| \neq |\mu_B|$. Note, however, that since $\mu_B = -\mu_A$, such a centering can be easily accomplished by appropriately scaling the stimulus space.

References

- Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision*, 2(1):5, 66–78, <http://journalofvision.org/2/1/5/>, doi:10.1167/2.1.5. [PubMed] [Article]
- Abbey, C. K., Eckstein, M. P., & Bochud, F. O. (1999). Estimation of human-observer templates for 2 alternative forced choice tasks. *Proceedings of SPIE*, 3663, 284–295.
- Ahumada, A. J. (1967). Detection of tones masked by noise: A comparison of human observers with digital-

- computer-simulated energy detectors of varying bandwidths. Unpublished doctoral dissertation, University of California, Los Angeles.
- Ahumada, A. J. (1996). Perceptual classification images from Vernier acuity masked by noise. *Perception*, *25*, 18.
- Ahumada, A. J., Jr. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, *2*(1):8, 121–131, <http://journalofvision.org/2/1/8/>, doi:10.1167/2.1.8. [PubMed] [Article]
- Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on inconsistencies between visual and haptic percepts. *Vision Research*, *41*, 449–461. [PubMed]
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *20*, 1391–1397. [PubMed]
- Beard, B. L., & Ahumada, A. J., Jr. (1999). Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *16*, 755–763. [PubMed]
- Bejjanki, V. R., Ma, W. J., Beck, J. M., & Pouget, A. (2007). Perceptual learning as improved Bayesian inference in early sensory areas. Poster presented at Computational and Systems Neuroscience Conference (CoSyNe), Salt Lake City, UT.
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9):1, 659–667, <http://journalofvision.org/5/9/1/>, doi:10.1167/5.9.1. [PubMed] [Article]
- Das, A. (1997). Plasticity in adult sensory cortex: A review. *Network*, *8*, R33–R76.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. [PubMed]
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*, 69–73. [PubMed] [Article]
- Fahle, M., & Poggio, T. (2002). *Perceptual learning*. Cambridge, MA: MIT Press.
- Geisler, W. S. (2003). Ideal observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 825–837). Boston: MIT Press.
- Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, *5*(11):7, 1013–1023, <http://journalofvision.org/5/11/7/>, doi:10.1167/5.11.7. [PubMed] [Article]
- Gibson, E. J. (1953). Improvement in perceptual judgments as a function of controlled practice or training. *Psychological Bulletin*, *50*, 401–431.
- Gilbert, C. D. (1994). Early perceptual learning. *Proceedures of the National Academy of Sciences United States of America*, *91*, 1195–1197. [PubMed] [Article]
- Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, *28*, 167–207.
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between senses. *Science*, *298*, 1627–1630. [PubMed]
- Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*, 3621–3629. [PubMed]
- Jacobs, R. A., & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research*, *39*, 4062–4075. [PubMed]
- Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research*, *43*, 831–854. [PubMed]
- Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as bayesian inference*. Cambridge: Cambridge University Press.
- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*, 2539–2558. [PubMed]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412. [PubMed]
- Levi, D. M., & Klein, S. A. (2002). Classification images for detection and position discrimination in the fovea and parafovea. *Journal of Vision*, *2*(1):4, 46–65, <http://journalofvision.org/2/1/4/>, doi:10.1167/2.1.4. [PubMed] [Article]
- Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision ‘template’ for position discrimination. *Nature Neuroscience*, *7*, 178–183. [PubMed]
- Lu, H., & Liu, Z. (2006). Computing dynamic classification images from correlation maps. *Journal of Vision*, *6*(4):12, 475–483, <http://journalofvision.org/6/4/12/>, doi:10.1167/6.4.12. [PubMed] [Article]
- Lu, Z.-L., & Doshier, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise [Special issue]. *Journal of the Optical Society of America A*, *16*, 764–778.

- Mannos, J., & Sakrison, D. (1974). The effects of a visual fidelity criterion on the encoding of images. *IEEE Transactions on Information Theory*, *20*, 4, 525–536.
- Marr, D. (1982). *Vision*. New York: W. H. Freeman and Company.
- Murray, R. F. (2002). Perceptual organization and the efficiency of shape discrimination. Unpublished doctoral dissertation, University of Toronto, Canada.
- Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, *5*, 812–816. [[PubMed](#)] [[Article](#)]
- Olman, C., & Kersten, D. (2004). Classification objects, ideal observers, and generative models. *Cognitive Science*, *28*, 227–240.
- Otto, T. U., Herzog M. H., Fahle, M., & Zhaoping, L. (2006). Perceptual learning with spatial uncertainties. *Vision Research*, *46*, 3223–3233. [[PubMed](#)]
- Petrov, A. A., Doshier, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, *112*, 715–743. [[PubMed](#)]
- Recanzone, G. H., Merzenich, M. M., & Jenkins, W. M. (1992). Frequency discrimination training engaging a restricted skin surface results in an emergence of a cutaneous response zone in cortical area 3a. *Journal of Neurophysiology*, *67*, 1057–1070. [[PubMed](#)]
- Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience*, *13*, 87–103. [[PubMed](#)] [[Article](#)]
- Simoncelli, E. P., Paninski, L., Pillow, J., & Schwartz, O. (2004). Characterizing neural responses with stochastic stimuli. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed., pp. 327–338). Boston: MIT Press.
- Teich, A. F., & Qian, N. (2003). Learning and adaptation in a recurrent model of V1 orientation selectivity. *Journal of Neurophysiology*, *89*, 2086–2100. [[PubMed](#)] [[Article](#)]
- Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge: Cambridge University Press.
- Zhaoping, L., Herzog, M. H., & Dayan, P. (2003). Nonlinear ideal observation and recurrent processing in perceptual learning. *Network*, *14*, 223–247. [[PubMed](#)]