

# Adaptive precision pooling of model neuron activities predicts the efficiency of human visual learning

Center for Visual Science, Department of Brain and  
Cognitive Sciences, University of Rochester,  
Rochester, NY, USA



**Robert A. Jacobs**

When performing a perceptual task, precision pooling occurs when an organism's decisions are based on the activities of a small set of highly informative neurons. The Adaptive Precision Pooling Hypothesis links perceptual learning and decision making by stating that improvements in performance occur when an organism starts to base its decisions on the responses of neurons that are more informative for a task than the responses that the organism had previously used. We trained human subjects on a visual slant discrimination task and found their performances to be suboptimal relative to an ideal probabilistic observer. Why were subjects suboptimal learners? Our computer simulation results suggest a possible explanation, namely that there are few neurons providing highly reliable information for the perceptual task, and that learning involves searching for these rare, informative neurons during the course of training. This explanation can account for several characteristics of human visual learning, including the fact that people often show large differences in their learning performances with some individuals showing no performance improvements, other individuals showing gradual improvements during the course of training, and still others showing abrupt improvements. The approach described here potentially provides a unifying framework for several theories of perceptual learning including theories stating that learning is due to adaptations of the weightings of read-out connections from early visual representations, external noise filtering or internal noise reduction, increases in the efficiency with which learners encode task-relevant information, and attentional selection of specific neural populations which should undergo adaptation.

Keywords: computational modeling, learning, depth

Citation: Jacobs, R. A. (2009). Adaptive precision pooling of model neuron activities predicts the efficiency of human visual learning. *Journal of Vision*, 9(4):22, 1–15, <http://journalofvision.org/9/4/22/>, doi:10.1167/9.4.22.

## Introduction

Are human visual judgments based on the activities of all neurons sensitive to a stimulus or just the activities of a specialized subgroup? If the latter, can human visual learning be characterized as an improvement in the selection of neurons to include in this subgroup?

A global pooling hypothesis states that the activities of all active neurons (or perhaps all active neurons in a particular brain region) contribute to an organism's percept and, thus, to an organism's perceptual decision (e.g., Shadlen, Britten, Newsome, & Movshon, 1996). A precision pooling hypothesis states that an organism's decision is determined solely by those neurons whose activities are believed to be highly informative about the correct answer (e.g., Purushothaman & Bradley, 2005). Consider, for example, a monkey discriminating two directions of visual motion in the presence of noise (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996). According to a global pooling hypothesis, the monkey reaches its decision by indiscriminately pooling the responses of a large population of neurons in, for instance, the middle temporal (MT) cortical area, an area specialized for visual motion analysis. Importantly, this population includes neurons with widely varying properties, such as different motion direction sensitivities (Nadler &

DeAngelis, 2005). In contrast, a precision pooling hypothesis states that the monkey reaches its decision by pooling the responses of just those neurons whose responses carry highly reliable information about the motion direction. In many cases, this population is a small subset of MT neurons (Jazayeri & Movshon, 2007; Nadler & DeAngelis, 2005; Purushothaman & Bradley, 2005).

Although a matter of current debate, there is now data consistent with a precision pooling hypothesis, at least under some circumstances. For example, single neurons have been found to be reliable detectors of perceptual stimuli (Barlow, 1995; Bradley, Skottun, Ohzawa, Sclar, & Freeman, 1987; Ohzawa, DeAngelis, & Freeman, 1990), the activities of individual neurons have been related to an organism's perceptual decisions (Parker & Newsome, 1998), and an organism's decisions have been found to be correlated with the activities of higher precision neurons but not with those of lower precision neurons (Purushothaman & Bradley, 2005). If the precision pooling hypothesis is at least partially correct, then this raises the possibility that there is a link between perceptual learning and the pooling of neural responses. Based on behavioral experiments, it has previously been hypothesized that improvements in performance occur when an organism starts to use perceptual features that are more informative for a task than the features that the organism had previously used (Gibson & Gibson, 1955).

To link perceptual learning and the pooling of neural responses, it is useful to restate this behavioral hypothesis at a neural level. According to this restatement, referred to as the Adaptive Precision Pooling (APP) Hypothesis, improvements in performance occur when an organism starts to base its decisions on the responses of neurons that are more informative for a task than the responses that the organism had previously used.

The APP Hypothesis has its roots in theories stating that visual learning effects are due to reweightings of the “read-out” connections from early visual representations (Doshier & Lu, 1998, 1999; Law & Gold, 2008; Petrov, Doshier, & Lu, 2005; Saarinen & Levi, 1995).<sup>1</sup> A contribution of this article is its examination of an extreme form of this idea in which the weightings are biased to be sparse, meaning that only a relatively small number of weights are nonzero, and thus, only a small subset of neurons contribute to decision making (see also Liu & Weinshall, 2000).

We trained human subjects on a visual slant discrimination task and measured their learning performances. We also computed the performances of a statistical model, referred to as an “ideal probabilistic observer”, on the same task (Barlow, 1957; Geisler, 1989; Green & Swets, 1966). The results show that subjects did not learn as well as the observer, meaning that they did not learn as much during training as they theoretically could have. Why were subjects suboptimal learners? Our simulation results suggest that the APP Hypothesis provides a possible explanation, namely that there are few neurons providing highly reliable information for the perceptual task, and learning involves searching for these rare, informative neurons during the course of training.

Although the APP Hypothesis potentially accounts for many characteristics of human visual learning, a key motivation for the work presented here is the need to account for differences in individuals’ learning performances. A common observation (or frustration) among scientists studying visual learning is that participants in experiments often show different dynamics in their performances with some individuals showing no performance improvements, other individuals showing gradual improvements during training, and still others showing abrupt improvements (Rubin, Nakayama, & Shapley, 2002). As illustrated below, the APP Hypothesis accounts for all three types of learning dynamics.

## Experiment

### Subjects

Eight naive subjects with normal or corrected-to-normal vision participated in the experiment. Subjects gave written informed consent before participation.

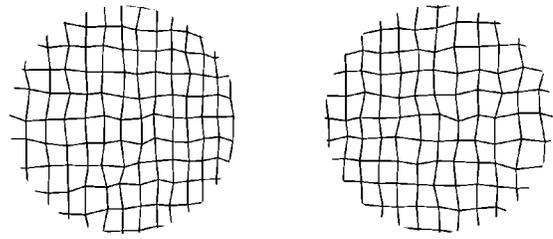


Figure 1. Illustrations of planar surfaces defined by “noisy” grids. The surfaces were viewed through a circular aperture so that their outer edges were not visible. The left surface has a slant of  $40^\circ$  (left side closer), whereas the right surface has a slant of  $-30^\circ$  (right side closer).

### Stimuli

Stimuli depicted planar surfaces visually defined by “noisy” grids of horizontal and vertical lines (see Figure 1). A noisy grid was created in two stages. In the first stage, a noise-free grid was placed on the surface when the surface was at a frontoparallel orientation. In the second stage, the spatial location of each of the points where the horizontal and vertical lines intersect was perturbed by a random value sampled from an isotropic two-dimensional Normal distribution. Surfaces were then rotated in depth about their vertical axes. The noisy grid created a visual texture pattern whose gradients of element density, area, and compression were (ambiguous) cues to a surface’s orientation in depth (Blake, Bühlhoff, & Sheinberg, 1993; Cumming, Johnston, & Parker, 1993; Cutting & Millard, 1984; Knill, 1998). Lastly, surfaces were placed behind a circular aperture so that their outer edges were not visible, thereby removing a contour cue to surfaces’ orientations. Displays were viewed at a viewing distance of approximately 65 cm. At this distance, a display subtended  $14.5^\circ$  of visual angle in the horizontal and vertical dimensions.

### Procedure

On each trial, a subject (monocularly) viewed a stimulus display for 500 ms, and then judged whether the display depicted a surface whose left side or right side was closer. The subject responded by hitting an appropriate key on a keyboard. On practice and training trials, a response was followed by an auditory sound indicating whether the response was correct or incorrect. On test trials, a response was followed by an auditory “click” sound indicating that the response was recorded by the computer.

Subjects participated for 3 experimental sessions where each session lasted approximately 50 minutes. On Day 1, subjects performed 40 practice trials in which surface slants were set to either  $-40^\circ$  or  $+40^\circ$ . This was followed by 240 test trials, 16 trials at each of 15 surface slants ( $0^\circ$ ,  $\pm 5^\circ$ ,  $\pm 10^\circ$ ,  $\pm 15^\circ$ ,  $\pm 20^\circ$ ,  $\pm 25^\circ$ ,  $\pm 30^\circ$ ,  $\pm 40^\circ$ ). Using maximum likelihood estimation, a psychometric function (a logistic

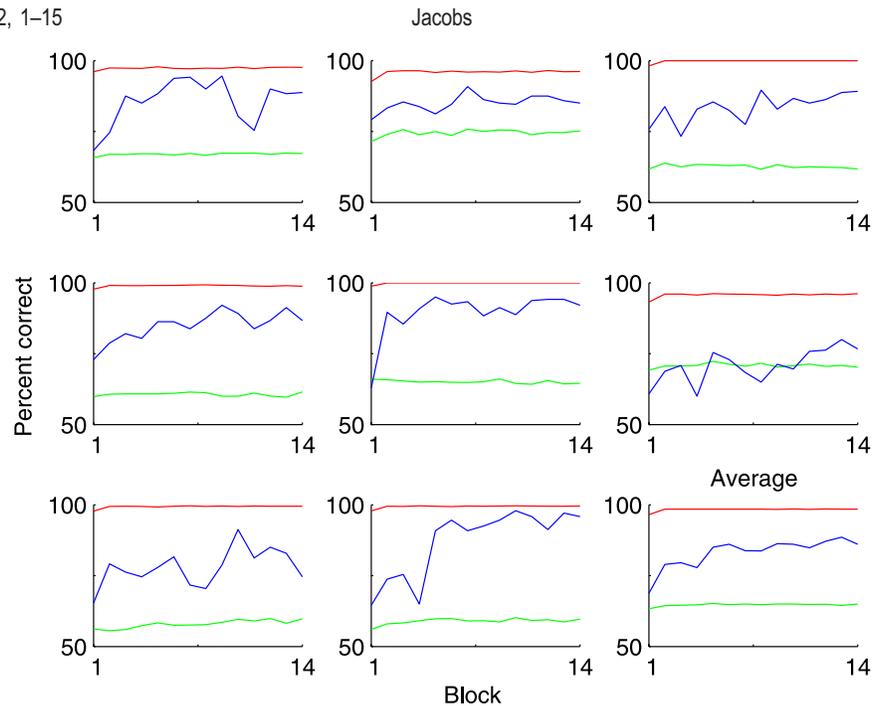


Figure 2. Learning curves for the subjects and the ideal probabilistic observers consistent with a global pooling hypothesis. The horizontal axis of each graph indicates the training block number, and the vertical axis indicates the percent correct. In each of the first 8 graphs, the blue line plots the performance of an individual subject, and the red and green lines plot the average performances of the Anisotropic (Gaussian Mixture Model with diagonal covariance matrix) and Isotropic (Gaussian Mixture Model with diagonal covariance matrix with identical values along the diagonal) Observers, respectively, when trained under the same conditions as the subject. The bottom right graph shows the averages of the first 8 graphs.

function) was fit to a subject's test data (the absolute value of a surface slant was the independent variable, and the probability that the subject responded correctly was the dependent variable). Based on this function, the slant value corresponding to a subject's 70%-correct threshold, denoted  $\tau$ , was calculated. The subject then performed 4 blocks of training trials in which each block consisted of 240 trials. Surface slants on training trials were either  $-\tau$  or  $+\tau$ . A subject performed 5 training blocks on each of Days 2 and 3.

## Results

Subjects' learning curves are shown in Figure 2 (blue lines). Importantly for our purposes, performances sometimes gradually improved and sometimes abruptly improved during the course of training. In addition, final performances were less than 100%. These results will be interpreted in the context of ideal probabilistic observers defined below.

## Computational simulations

Within the vision sciences, there is debate over whether visual learning effects are best characterized as resulting

from changes in early sensory representations (Karni & Sagi, 1991; Schoups, Vogels, Qian, & Orban, 2001) or as resulting from learned reweighting of the read-out connections from early visual representations with no changes in the representations themselves (Doshier & Lu, 1998, 1999; Law & Gold, 2008; Petrov et al., 2005; Saarinen & Levi, 1995). In this work, our approach to developing probabilistic observers for visual learning has been to develop models with a fixed visual representation but with adaptable read-out parameters governing models' judgments about visual stimuli. The input to our observers was the output of a fixed representational front end. The observers were adaptable Gaussian mixture models.

## Representational front end

The representational front end was a model neural circuit intended to mimic important aspects of primate visual cortex. We used the same front end (with the same parameter values) as Petrov et al. (2005; see their article for additional details). This front end can be regarded as an instance of a "back pocket model" (Chubb & Landy, 1991), a type of model commonly used by vision scientists to account for aspects of visual texture perception (Chubb & Talevich, 2002 list 23 articles in the scientific literature using a back pocket model). These models typically start with a stage of processing in which

a battery of spatially local, linear filters is applied to the visual input. The receptive fields of these filters resemble simple cell receptive fields of various spatial frequencies and orientations (e.g., Gabor filters). The models often also include nonlinear processing in the form of a pointwise nonlinearity (such as a rectifier) and in the form of lateral interactions between neural units (such as cross-orientation or cross-frequency normalization; Heeger, 1992), as well as pooling of local signals to create signals with more global properties.

The input to the front end was a pixel-based representation of images from our psychophysical task (i.e., each image depicted a planar surface with a noisy grid slanted in depth). An image was filtered by convolving it with 140 receptive fields, where each field was a Gabor function with orientation  $\theta \in \{0, \pm 15, \pm 30, \pm 45\}$ , spatial frequency  $f \in \{1, 1.4, 2, 2.8, 4\}$ , and phase  $\phi \in \{0, 90, 180, 270\}$ . The filtered images were then rectified using the half-squaring operator. The resulting values formed a set of phase-sensitive maps (one for each orientation, frequency, and phase), which can be interpreted as activation patterns across a retinotopic population of simple cells in area V1 (Heeger, 1992). The retinotopic maps of phase-sensitive units were then combined into phase-invariant maps (analogous to complex cells) by pooling phase-sensitive units that form quadrature pairs, as is done in energy models of visual motion perception (Adelson & Bergen, 1985). Each phase-invariant map was normalized by dividing it by a frequency-dependent normalization term, thereby producing a map whose total activation was approximately constant for above-threshold stimulus contrasts (Heeger, 1992). Lastly, each phase-invariant map was pooled across space using a Gaussian weighting kernel.

In our simulations, the displays used in the experiment were downsampled to create  $128 \times 128$  images. Each image was divided into 9 partially overlapping  $64 \times 64$  regions (regions were defined by dividing the original image into 3 [partially overlapping] rows and 3 [partially overlapping] columns). Spatial averaging took place within each region. The final output of the representational front end was the response values of 315 output units (9 regions  $\times$  5 frequencies  $\times$  7 orientations). These response values were the inputs to the probabilistic observers.

## Gaussian mixture models

Each ideal probabilistic observer was a Gaussian mixture model with two Gaussian distributions, one for images depicting left-closer surfaces and the other for images depicting right-closer surfaces. On each trial of the slant discrimination task, an observer decided whether an image depicted a left-closer or right-closer surface. This decision was made as follows. Let  $\vec{x}_t$  denote the values of the output units of the representational front end on trial  $t$

(because the front-end outputs are the inputs to a model,  $\vec{x}_t$  also denotes these inputs). In addition, let  $y_t^*$  be a binary variable indicating whether the image depicted a left-closer ( $y_t^* = 1$ ) or right-closer surface ( $y_t^* = 0$ ) on trial  $t$ . Finally, let  $\vec{w}_t$  denote the observer's parameter values (mean vectors and covariance matrices for the left-closer and right-closer distributions) on trial  $t$ . The observer computed the posterior probability that the image depicted a left-closer surface via Bayes' rule:

$$P(y_t^* = 1 | \vec{x}_t, \vec{w}_t) = \frac{P(\vec{x}_t | y_t^* = 1, \vec{w}_t)P(y_t^* = 1)}{P(\vec{x}_t | y_t^* = 1, \vec{w}_t)P(y_t^* = 1) + P(\vec{x}_t | y_t^* = 0, \vec{w}_t)P(y_t^* = 0)}, \quad (1)$$

where the prior probabilities  $P(y_t^* = 1)$  and  $P(y_t^* = 0)$  were each set to 1/2. If  $P(y_t^* = 1 | \vec{x}_t, \vec{w}_t) \geq 0.5$ , then the observer decided "left-closer"; otherwise, it decided "right-closer."

The probabilistic observer adapted its parameter values on each trial. These parameter values always equaled the best possible values, and thus, our probabilistic observers can also be regarded as ideal learners (Eckstein, Abbey, Pham, & Shimozaki, 2004; Michel & Jacobs, 2008). On trial  $t$ , the observer had viewed the images on all trials up to and including trial  $t$ , along with the corresponding correct judgments of these images. The observer should use all of this information to optimally set its parameters. Let  $X_{1,\dots,t}$  denote the values of the front-end output units on trials 1 through  $t$  (i.e.,  $\vec{x}_1, \dots, \vec{x}_t$ ), and let  $Y_{1,\dots,t}^*$  denote the corresponding values of the binary target variable (i.e.,  $y_1^*, \dots, y_t^*$ ). The optimal setting of the parameter values (in a maximum a posteriori sense) is the values that maximize the posterior distribution of the parameter values given the data  $X_{1,\dots,t}$  and  $Y_{1,\dots,t}^*$ . Using Bayes' rule, this distribution can be written as

$$P(\vec{w}_{t+1} | X_{1,\dots,t}, Y_{1,\dots,t}^*) \propto P(Y_{1,\dots,t}^* | X_{1,\dots,t}, \vec{w}_{t+1})P(\vec{w}_{t+1}). \quad (2)$$

The second term on the right-hand side is the prior distribution of parameter values. We assume that this is a uniform distribution. The first term is the likelihood function. Assuming that the data on each trial are independent, the likelihood function can be rewritten as

$$P(Y_{1,\dots,t}^* | X_{1,\dots,t}, \vec{w}_{t+1}) = \prod_{i=1}^t P(y_i^* | \vec{x}_i, \vec{w}_{t+1}). \quad (3)$$

An ideal probabilistic observer maximized this likelihood function as follows. Let  $I_L$  be the set of trial indices for trials in which a left-closer surface was depicted, and let  $I_R$  be the set of trial indices for trials in which a right-

closer surface was depicted.  $I_L$  and  $I_R$  are disjoint sets whose union is the set  $\{1, \dots, t\}$ . The mean vector of an observer’s left-closer Gaussian distribution was set to the mean of the data  $\{\bar{x}_i\}_{i \in I_L}$ , whereas the mean vector of the right-closer distribution was set to the mean of the data  $\{\bar{x}_i\}_{i \in I_R}$ .

In the simulations described below, the covariance matrices of the left-closer and right-closer distributions were restricted so as to reduce the number of parameters whose values needed to be estimated, to make the estimation of the parameter values easy and robust, and to make the simulation results easy to analyze. Let the matrices  $\Sigma_L$  and  $\Sigma_R$  denote the covariances of the data  $\{\bar{x}_i\}_{i \in I_L}$  and  $\{\bar{x}_i\}_{i \in I_R}$ , respectively. In some simulations (those of the Anisotropic Observer described below),  $\Sigma_L$  and  $\Sigma_R$  were restricted to be diagonal matrices, meaning they only represented the variances of the inputs, and covariances among pairs of inputs were assumed to be zero. In other simulations (those of the Isotropic Observer described below), each of these matrices was further restricted to have the same value along the diagonal, equal to the average variance of the inputs. Lastly, our simulations assumed that the left-closer and right-closer distributions had equal covariance matrices. This common matrix, denoted  $\Sigma$ , was set to the pooled covariance matrix:

$$\Sigma = \frac{n_L}{n_L + n_R} \Sigma_L + \frac{n_R}{n_L + n_R} \Sigma_R, \quad (4)$$

where  $n_L$  and  $n_R$  are the number of elements in  $I_L$  and  $I_R$ , respectively. In the field of statistics, this pooled covariance matrix is also referred to as the average class-conditional covariance matrix.

## Simulation results

We simulated two sets of probabilistic observers. One set is consistent with a global pooling hypothesis in the sense that its members received the responses of all front-end output units. The second set is consistent with a precision pooling hypothesis in the sense that its members received the responses of a subset of front-end output units.

The first set consisted of two observers referred to as the Anisotropic and Isotropic Observers. For the Anisotropic Observer, its two Gaussian distributions shared a common covariance matrix, and this matrix was restricted to a diagonal matrix. The term “anisotropic” refers to the fact that each front-end output unit (or, equivalently, each input to the observer) was modeled with its own variance parameter, meaning that the Anisotropic Observer scaled its input space differently in different input dimensions.<sup>2</sup> This observer had 945 parameters: 315 mean parameters for each distribution and 315 variance parameters shared by both distributions.

The Isotropic Observer was identical to the Anisotropic Observer except that its covariance matrix was further restricted such that its diagonal entries were equal to each other. Because the same variance parameter was used for all inputs, the Isotropic Observer used a uniform scaling in all input dimensions. This observer had 631 parameters: 315 mean parameters for each distribution and 1 shared variance parameter.

For each subject in the psychophysical experiment, 25 simulations of the Anisotropic and Isotropic Observers were performed. Each simulation used the same sequence of slants as the subject. That is, if the subject viewed a surface with slant  $\tau$  on training trial  $t$ , then an observer was exposed to an image depicting a surface with slant  $\tau$  on its  $t$ th training trial. The replications of each observer differed in the noise added to the grids defining each surface (see Figure 1).

We computed the Anisotropic and Isotropic Observers’ average learning curves (see Figure 2). A notable feature of these curves is that they are relatively flat, i.e., the observers’ learning algorithms converged quickly and their performances did not significantly improve beyond the first training block. In this sense, the human subjects’ learning curves and the observers’ learning curves have different shapes suggesting that their underlying learning mechanisms have different dynamics. This important point will be discussed below in the [Adaptive precision pooling hypothesis](#) section.

The Anisotropic Observer performed nearly perfectly starting from the first training block (red lines in Figure 2). This result indicates that, in the 315-dimensional space defined by the responses of the front-end output units, the left-closer and right-closer surfaces are far apart and, thus, easy for the Anisotropic Observer to distinguish. In contrast, the Isotropic Observer showed relatively poor performance (green lines in Figure 2). A comparison of the Anisotropic and Isotropic Observers highlights the importance of individually scaling each front-end output unit’s response. When each unit’s response was individually scaled, as in the Anisotropic Observer, the visual slant discrimination task is easy. When the same scaling is applied to all units’ responses, as in the Isotropic Observer, the task is difficult.

The results also indicate that our human subjects were inefficient learners (inefficiency of human visual learning has previously been reported by Eckstein et al., 2004 and Michel & Jacobs, 2008). Whereas the Anisotropic Observer achieved near-perfect performance starting from the first training block, none of our subjects achieved near-perfect performance after fourteen blocks of training. An observer’s discrimination performance can be quantified in units of  $d'$ , a measure based on signal detection theory (Green & Swets, 1966). If we let  $d'_{\text{subj}}(b)$  denote a subject’s performance on training block  $b$ , and let  $d'_{\text{AO}}(b)$  denote the Anisotropic Observer’s performance on training block  $b$ , then  $(d'_{\text{subj}}(b)/d'_{\text{AO}}(b))^2$  is the subject’s discrimination efficiency at block  $b$  (this quantity goes

between 0 and 1; a value of 1 means that the subject performed optimally in the sense that the subject performed as well as the ideal probabilistic observer, whereas a value less than 1 indicates that the subject performed suboptimally; Tanner & Birdsall, 1958). On average, subjects' discrimination efficiencies were 0.09 on block 1 and increased to 0.31 on block 14. These data indicate that subjects did not learn as much as they theoretically could have (based on the assumptions underlying the Anisotropic Observer).

Above, we stated that the Anisotropic Observer is consistent with a global pooling hypothesis because it receives the responses of all front-end output units. Although it receives the responses of all units, this does not necessarily mean that it makes significant use of all responses. It is possible that this observer uses the responses of a subset of the units and ignores all other units. We evaluated this possibility as follows. The Anisotropic Observer is an instance of a Gaussian Mixture Model with two components, and thus, it is possible to convert it to an equivalent logistic regressor (Bishop, 2006). This regressor maps the activities of the front-end output units to the posterior probability that an image depicts a surface whose left side is closer to the observer (one minus this value is the probability that a surface's right side is closer). Let  $\bar{x}_i^L(t)$  denote the mean of the  $i$ th unit's activities over trials 1, ...,  $t$  when images depicted left-closer surfaces, and let  $\bar{x}_i^R(t)$  denote the corresponding

value when images depicted right-closer surfaces. Let  $\sigma_i^2(t)$  denote the unit's average class-conditional variance. The regression coefficient corresponding to the  $i$ th front-end output unit at training trial  $t$  is  $w_i(t) = (\bar{x}_i^L(t) - \bar{x}_i^R(t)) / \sigma_i^2(t)$ . To quantify the degree to which the regressor makes use of unit  $i$  at trial  $t$ , we define the relative weight for unit  $i$  as  $r_i(t) = |w_i(t)| / \sum_j |w_j(t)|$ .

Based on their average relative weights at the end of training, we rank ordered the front-end output units. The results are that 10%, 20%, and 30% of the units account for 44%, 67%, and 82% of the regressor's total relative weight, respectively. To account for 95% of the regressor's total relative weight requires 46% of the units. In summary, the logistic regressor derived from the Anisotropic Observer makes nearly no use of about half of the front-end output units and makes at least mild use of the remaining units. We conclude that, although the Anisotropic Observer has the potential to follow the decision making strategy postulated by a global pooling hypothesis, it makes nonnegligible use of only half of its inputs, which means that the observer does not closely follow this strategy.

The second set of ideal probabilistic observers consists of observers that are consistent with a precision pooling hypothesis. Each observer was identical to the Anisotropic Observer with the exception that it received the responses of only a subset of front-end output units. The size of the subset was fixed, but the units comprising the subset were

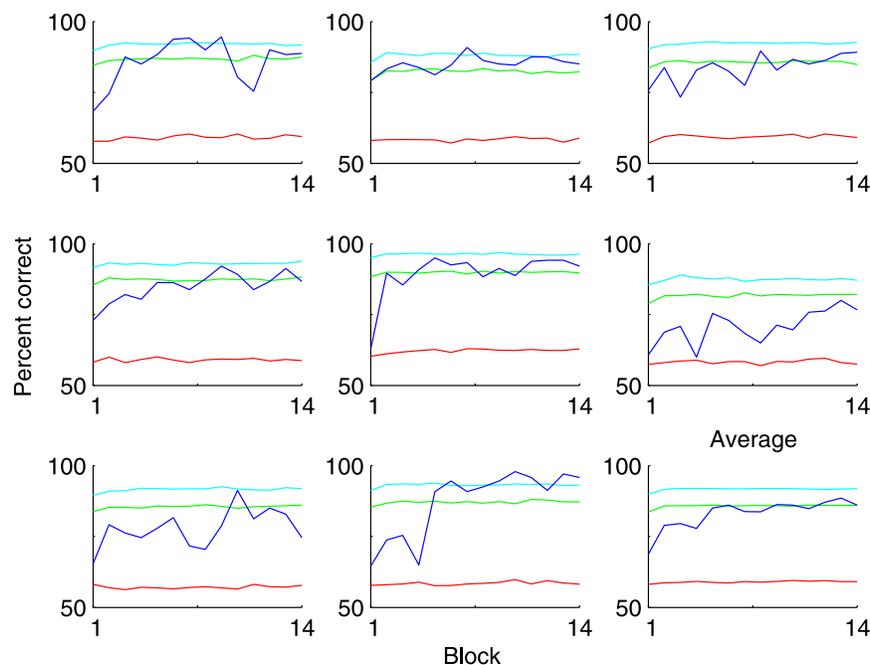


Figure 3. Learning curves for the subjects and the ideal probabilistic observers consistent with a precision pooling hypothesis. The horizontal axis of each graph indicates the training block number, and the vertical axis indicates the percent correct. In each of the first 8 graphs, the blue line plots the performance of an individual subject, and the red, green, and cyan lines plot the average performances of the observers that used 1, 8, or 16 front-end output units chosen at random, respectively, when trained under the same conditions as the subject. The bottom right graph shows the averages of the first 8 graphs.

chosen at random. For each subject in the psychophysical experiment, 25 simulations of a probabilistic observer were performed. The replications of an observer differed in the noise added to the grids defining a surface shown on a trial and in the subset of front-end output units whose activities were the inputs to the observer.

The data are shown in Figure 3. As before, the observers' learning curves are flat indicating that their learning dynamics were different than the dynamics characterizing learning in the human subjects. Again, this point will be addressed below.

An observer that used only a single front-end output unit performed poorly on average on the visual slant discrimination task (red line in Figure 3). An observer that used 8 units (2.5% of the total number of front-end output units) performed well (green line), and an observer that used 16 units (5%) performed extremely well (cyan line). Subjects' performances (blue line) at the end of training were nearly equal to those of the observer that used 8 units. These data reveal that there is a "point of diminishing returns" when considering how many units to use during decision making. For example, the Anisotropic Observer used many more units (315) than the observer that used 16 units, but its performance was only mildly better.

## Adaptive precision pooling hypothesis

Intuitively, it seems that a decision network computing on the basis of a global pooling strategy would be inefficient because this network would include in its calculations neural responses that are irrelevant for the current task, responses that are highly noisy, or both. Thus, it seems sensible that a neural decision network might be biased toward the use of only a limited subset of inputs. Speaking broadly, this intuition has been formalized previously in different ways by both computational scientists and neurobiologists.

First, a decision network consistent with global pooling could perform its calculations on the basis of the activities of hundreds of thousands of neurons. However, it has been found that computational techniques that work well with low-dimensional input spaces often perform poorly with high-dimensional input spaces, a problem known as the "curse of dimensionality" (Bellman, 1957). Limiting the inputs to a network, as in precision pooling, can potentially be a form of dimensionality reduction, which ameliorates this problem. Second, a decision network receiving the inputs of many neurons has many parameters it could adapt (e.g., synaptic strengths). In the computational literature, it has been found that methods with many parameters are prone to "over-fitting" their training data, thereby leading to poor generalization. A

bias toward the use of only a limited subset of inputs can be regarded as a type of regularization, which may lead to better task performance with novel stimuli (MacKay, 2003). Third, computational neuroscientists have argued that sparse representations in which few neurons are concurrently active have important representational advantages (Földiák, 1990; Olshausen & Field, 1996; Simoncelli & Olshausen, 2001). For example, neurons in networks biased toward the use of only a limited set of inputs become sensitive to additive components, often referred to as independent components, underlying non-Gaussian source signals. Lastly, visual neuroscientists have argued that the high metabolic energy costs of spikes motivate both the use of representational codes that rely on few active neurons and the need for mechanisms of selective attention (Lennie, 2003).

Neurophysiological data reviewed above and elsewhere (Purushothaman & Bradley, 2005) suggest that biological organisms may make perceptual decisions using precision pooling under some circumstances. Our computer simulation results reported above suggest that precision pooling can be an effective decision making strategy showing near-optimal performance with relatively few computational resources. The arguments presented in the preceding paragraph further motivate the study of precision pooling. Despite the potential benefits of precision pooling, it also has significant shortcomings. Importantly for our purposes, it does not provide an adequate account of the learning dynamics and performances of the human subjects in our psychophysical experiment. This failure stems from the fact that, by itself, precision pooling does not provide a strong link between perceptual learning and decision making. In the remainder of this article, we examine the Adaptive Precision Pooling Hypothesis, which links learning and decision making.

The inputs to our ideal probabilistic observers were the activities of the model neurons, which serve as the representational front-end outputs. Therefore, the observers made their perceptual judgments on the basis of these neurons' activities. We examined the information carried by each of these model neurons as follows. As above, let  $\bar{x}_i^L(t)$  and  $\bar{x}_i^R(t)$  denote the means of the  $i$ th neuron's activities over trials 1, ...,  $t$  when images depicted left-closer and right-closer surfaces, respectively, and let  $\sigma_i(t)$  denote the neuron's average class-conditional standard deviation. The information carried by model neuron  $i$  at trial  $t$  can be quantified as  $d'_i(t) = |\bar{x}_i^L(t) - \bar{x}_i^R(t)| / \sigma_i(t)$  (Green & Swets, 1966). If neuron  $i$ 's activities are very different when images depict left-closer versus right-closer surfaces, then  $d'_i$  will be a large number. Otherwise, it will be near zero.

For each subject in the psychophysical experiment, we performed 25 simulations of the Anisotropic Observer. We then rank ordered the model neurons based on their average  $d'$  values at the end of training. The results are shown in Figure 4. It is consistently the case that a small number of neurons are highly informative for the visual

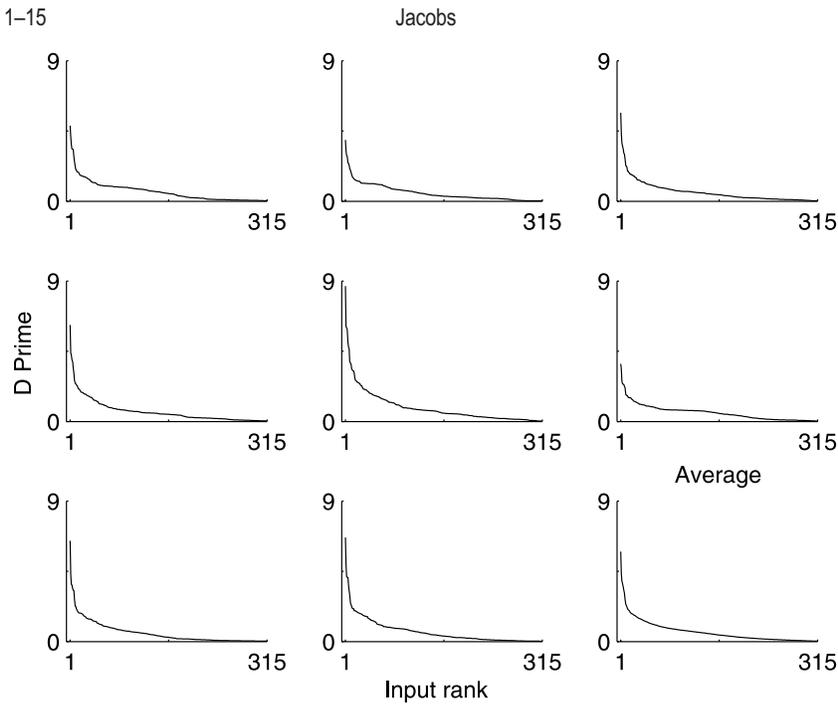


Figure 4. For each human subject, the Anisotropic Observer was simulated 25 times using the same training conditions as the subject. The model neurons that served as input to the observer were rank ordered based on their average  $d'$  values at the end of training. The horizontal axis of each graph plots a neuron's rank order, and the vertical axis plots a neuron's average  $d'$ . The first 8 graphs correspond to the simulations associated with the 8 subjects. The bottom right graph is the average of the first 8 graphs.

slant discrimination task, whereas the vast majority of neurons are only moderately or mildly informative.<sup>3</sup>

Motivated by these data, we begin to evaluate the APP Hypothesis by examining an observer that represents an extreme instance of a decision maker consistent with this hypothesis. As with previous observers, this observer was a Gaussian Mixture Model with two Gaussian distributions. Importantly, this observer performed the visual slant discrimination task on the basis of the responses of a single neuron in which the selection of this neuron was performed as follows. For each experimental subject, we rank ordered the model neurons according to their average  $d'(t)$  values at the end of training. We then simulated observers based on neurons ranked 1, 5, 10, and 20. The results are shown in Figure 5. When averaged across subjects, observers based on neurons ranked 1, 5, 10, and 20 achieved performances of 99%, 95%, 87%, and 80% correct at the end of training, respectively. Surprisingly, to achieve approximately the same level of performance as the human subjects, an observer needed to use the 10th most informative neuron (out of the population of 315 neurons).

The fact that an observer needed to use a model neuron in the top 3% of most informative neurons to match human performance suggests an explanation as to why human subjects in our experiment were suboptimal learners. It may be that, during the course of training, subjects “searched” for the neurons in visual cortex which were most informative for the experimental task.

When a subject found a neuron that was more informative than the neurons it had previously identified, the subject's performance improved. Because highly informative neurons are rare, a subject's performance was frequently imperfect.

To evaluate this explanation, and to further evaluate the APP Hypothesis, we implemented an observer that is an adaptive decision maker consistent with this hypothesis. This observer used a competitive learning scheme to perform a greedy hill-climbing search for a good model neuron. Consequently, this scheme closely resembles the model using competitive learning and constrained computational capacity previously proposed by Liu and Weinshall (2000) to account for aspects of generalization during human perceptual learning. At the start of a simulation, two model neurons were selected at random. One of these was randomly selected as the “winner,” whereas the other is referred to as the “competitor.” During a training block, the observer's judgments were based on the winning neuron. However, judgments based on the competitor were also calculated. At the end of a training block, the neuron with the best performance was designated as the new winner. The other neuron was discarded, and a new model neuron was randomly sampled to serve as the competitor. This learning process continued for the duration of training.

The results are shown in Figure 6 where each graph plots an observer's learning curve. In some simulations, performance improved gradually with training time. In

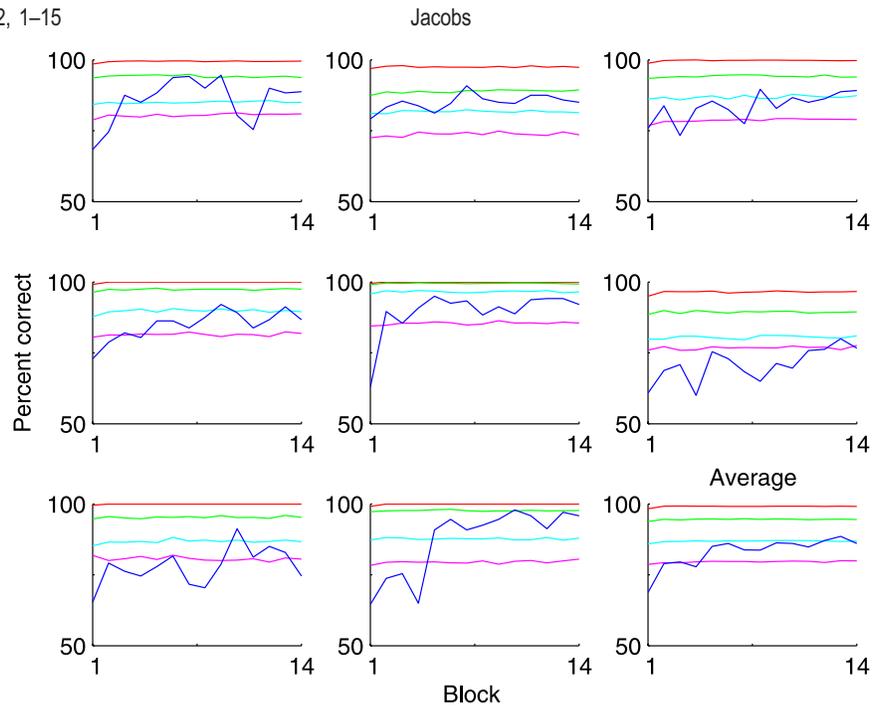


Figure 5. For each human subject, an ideal probabilistic observer was simulated 25 times using the same training conditions as the subject. This observer performed the visual slant discrimination task on the basis of the responses of a single neuron. The model neurons that served as input to the observers were rank ordered based on their average  $d'$  values at the end of training. The red, green, cyan, and magenta lines show the performances of the observers based on neurons ranked 1, 5, 10, and 20, respectively. The blue lines show the performances of the human subjects. The first 8 graphs plot the data associated with the 8 subjects. The bottom right graph is the average of the first 8 graphs.

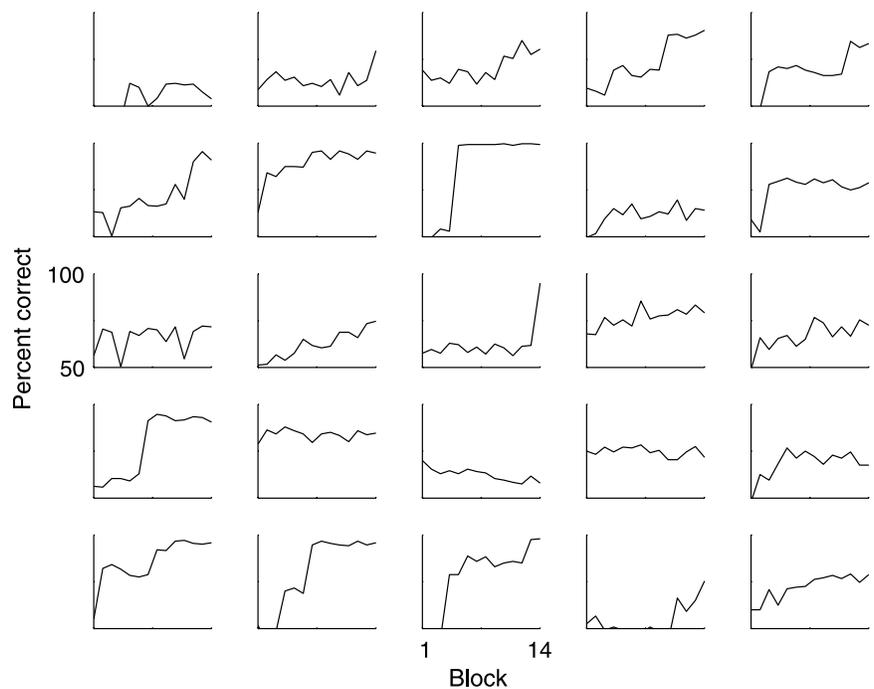


Figure 6. An observer implementing an adaptive precision pooling strategy was simulated 25 times using the same training conditions as a typical subject (Subject 7). Within a training block, this observer performed the visual slant discrimination task on the basis of the responses of a single neuron. Over training blocks, the observer performed a greedy hill-climbing search to identify a neuron whose responses were highly informative for the experimental task.

other simulations, performance rose abruptly, sometimes to a near-optimal level. In still other simulations, performance improvements were small or nonexistent. In terms of their shape and scale, these learning curves qualitatively resemble those of the human subjects. That is, the learning process studied here produces a range of learning performance dynamics that resembles the range of dynamics exhibited by human subjects.

To highlight the learning dynamics of the observer consistent with the APP Hypothesis, we contrast them with those of an observer that implemented an adaptive global pooling strategy. This observer performed the slant discrimination task using the responses of all front-end output neurons. On each trial, it updated its mean vectors and diagonal covariance matrix using conventional online error-correcting rules whose general forms are commonplace in the computational modeling literature (e.g., Petrov et al., 2005; Rescorla & Wagner, 1972; Widrow & Hoff, 1960). To update its mean vectors, it used the following rule:

$$\vec{\mu}_{\{L,R\}}(t+1) = \vec{\mu}_{\{L,R\}}(t) + \varepsilon[\vec{x}(t) - \vec{\mu}_{\{L,R\}}(t)], \quad (5)$$

where  $t$  indexes the trial number,  $\varepsilon$  is a learning rate parameter,  $\vec{x}(t)$  is a vector of front-end output neuron responses on trial  $t$ , and  $\vec{\mu}_{\{L,R\}}(t)$  is the mean vector of either the left-closer or right-closer Gaussian distribution, depending on which type of visual stimulus was displayed

on trial  $t$ . The update rule for the observer's covariance matrix was

$$\Sigma(t+1) = \Sigma(t) + \varepsilon[\mathbf{diag}[(\vec{x}(t) - \vec{\mu}_{\{L,R\}}(t))^2] - \Sigma(t)], \quad (6)$$

where the **diag** operator turns a vector into a diagonal matrix, and where the square of a vector is performed on an element-by-element basis. The only free parameter is the learning rate  $\varepsilon$ , which was set to a value ( $\varepsilon = 0.001$ ) that produced levels of performance at the end of training approximately equal to those of the human subjects. The results are shown in Figure 7. In all simulations, the learning curves have a stereotyped pattern in which performance improved in a slow, steady manner. Although these learning curves resemble those obtained by averaging the curves of many subjects, they do not resemble those of individual subjects. That is, the learning process of the adaptive global pooling observer does not produce a range of learning performance dynamics matching the range of dynamics exhibited by individual human subjects.

The adaptive precision pooling model differs from other models in several respects. It differs from linear models (or quasi-linear models such as logistic regression), such as that of Petrov et al. (2005) as well as the adaptive global pooling observer, because it includes a nonlinear adaptation mechanism. This mechanism produces a

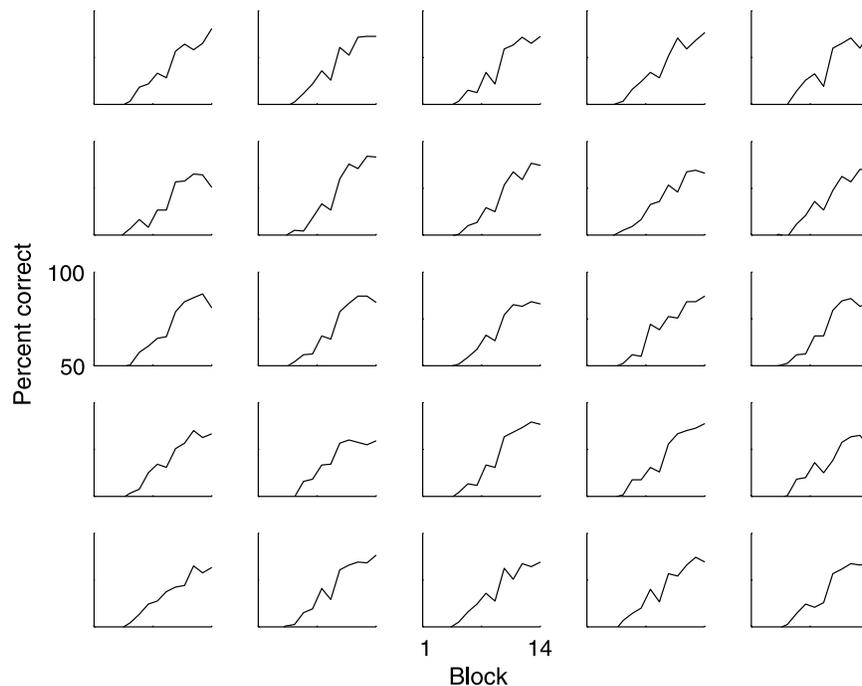


Figure 7. An observer implementing an adaptive global pooling strategy was simulated 25 times using the same training conditions as a typical subject (Subject 7). This observer performed the visual slant discrimination task on the basis of all front-end neuron responses. On each training trial, it used an online error-correcting rule to update its mean vectors and covariance matrix.

potential discontinuity in the model's behavior when it adds or deletes neurons from the limited set of neurons influencing decision making. This nonlinearity allows the model to exhibit a wide range of learning dynamics. In addition, the form of this nonlinearity distinguishes this model from other nonlinear models. For example, the model differs from multilayer neural networks, which learn new nonlinear features by adapting the weights of internal units (Rumelhart, Hinton, & Williams, 1986). While both forms of nonlinear adaptation are likely to play a role in human visual learning, this article has focused on the form exhibited by the adaptive precision pooling model. This focus is motivated, at least in part, on new neurophysiological evidence favoring precision pooling and on existing evidence regarding neural competition. Indeed, neural competition is thought to play a role in many forms of learning and development (Purves & Lichtman, 1985), and our simulations suggest that it also plays a role in the type of perceptual learning studied here.

## Discussion

We have described an account of human visual learning on a visual slant discrimination task, referred to as the Adaptive Precision Pooling (APP) Hypothesis. Relative to other possible accounts, we believe that this account has several appealing features.

First, the APP Hypothesis links behavior with its neural underpinnings. At a behavioral level, the hypothesis states that improvements in performance occur when a learner starts to use perceptual features that are more informative for a task than the features that the learner had previously used (Gibson & Gibson, 1955). At a neural level, the hypothesis states that an organism makes perceptual decisions by pooling the responses of a small number of highly informative neurons (Purushothaman & Bradley, 2005). In addition, a learner can adapt the set of neurons it uses during decision making to include more informative neurons. The behavioral consequence of this neural adaptation is that the learner's performance improves because it starts to use more informative features. In this article, we have considered the APP Hypothesis primarily from psychophysical and computational perspectives. We have not attempted to address its neurophysiological implementation. One possibility is that the selection of neurons to use during decision making can be performed through a competitive learning scheme based on Hebbian and anti-Hebbian mechanisms. Future work will need to explore this possibility. Future work will also need to explore more realistic models of neural function including models containing correlated noise among neural responses and models containing noise at the level of pooling and decision making.

Second, the APP Hypothesis potentially links theories of visual learning, attention, and decision making. We have examined a simple implementation of the hypothesis in this article. In particular, the simulations described above used a random hill-climbing search for potentially highly informative neurons. Our approach, however, provides the opportunity for new theories linking learning and attention based on mechanisms for selective attention which guide the search for informative neurons. We see this as an important area for future research. Early steps in this direction have already been taken (e.g., Dayan, Kakade, & Montague, 2000; Fahle, 2004; Hochstein & Ahissar, 2002).

Third, the vision sciences literature contains many theories of visual learning, including theories stating that learning is due to adaptations of low-level sensory representations (Karni & Sagi, 1991; Schoups et al., 2001), adaptations of the weightings of read-out connections from early visual representations (Doshier & Lu, 1998, 1999; Law & Gold, 2008; Petrov et al., 2005), external noise filtering or internal noise reduction (Doshier & Lu, 1998), increases in the efficiency with which learners encode task-relevant information (Gold, Bennett, & Sekuler, 1999), and attentional selection of specific neural populations which should undergo adaptation (Hochstein & Ahissar, 2002). The APP Hypothesis contains several of the best characteristics of many of these theories. The hypothesis clearly involves adaptations of the weightings of read-out connections from early visual representations. Because it also includes the addition of more informative neurons in the decision making process and the deletion of less informative neurons, a learner consistent with the hypothesis can increase the efficiency with which it encodes task-relevant information and can also perform external noise filtering and internal noise reduction. Furthermore, as mentioned above, the search for more informative neurons can be characterized as an attentional process guiding visual performance and learning. The key point here is that other theories in the vision sciences literature may be able to account for our human subjects' learning performances, but these theories are not necessarily competitors of the APP Hypothesis. Rather, the APP Hypothesis may provide a single mechanism for implementing these other theories. If so, it would provide a unifying framework for many current theories of perceptual learning.

Fourth, the APP Hypothesis provides new explanations for behavioral phenomenon which are currently poorly understood. It is often the case, for instance, that some human subjects in a visual learning experiment show gradual improvements in performance, whereas others show abrupt improvements (Rubin et al., 2002). The APP Hypothesis provides a single mechanism that can show both patterns of performance improvements (see Figure 6). As a second example, researchers have reported that perceptual learning is often highly specific. Subjects trained on a task at one retinal position, or with one eye,

or with one stimulus orientation may show improvements with practice, but their performances often return to baseline when tested under novel conditions (Fahle, Edelman, & Poggio, 1995; Fiorentini & Berardi, 1980; Karni & Sagi, 1991). According to the hypothesis, a learner will fail to generalize acquired perceptual knowledge when the neural responses selected by the learner's decision network are informative under training but not test conditions. Further experimentation is needed to evaluate these explanations.

Fifth, the APP Hypothesis has a solid statistical foundation. The implementation described here relies on statistical techniques which are well defined and well understood, such as the use of Gaussian Mixture Models for density estimation and the use of Bayes' rule for decision making (Bishop, 2006). In addition, the use of a greedy hill-climbing strategy to search for informative neurons resembles feature selection techniques that are commonly used in statistical applications when many features are available but only a small subset are likely to be relevant for a given task (Guyon & Elisseeff, 2003).

Sixth, the APP Hypothesis is simple. The implementation described here relies on low-order summary statistics such as means and variances. There are good reasons to believe that neurons could calculate these quantities. Indeed, computational neuroscientists have shown how neural circuits can encode and manipulate probability distributions (Pouget, Dayan, & Zemel, 2003).

Seventh, the APP Hypothesis suggests possible new research directions, including new experiments and predictions about the outcomes of those experiments. Consequently, the hypothesis is falsifiable. At a behavioral level, the hypothesis states that improvements in task performance are due to the use of new perceptual features. Vision scientists have recently begun to use behavioral techniques to determine the stimulus components, template, or "classification image" that an observer uses to perform a binary discrimination task (Abbey & Eckstein, 2002; Ahumada, 1967, 2002; Levi & Klein, 2002; Lu & Liu, 2006). In principal, and sometimes in practice, these techniques can be used to study how classification images change with learning (Beard & Ahumada, 1999; Gold, Sekuler, & Bennett, 2004; Michel & Jacobs, 2008). Consider an experiment in which stimuli are created from a large number of perceptual features, but only a small subset of these features are diagnostic of a stimulus category. One possibility, consistent with a global pooling hypothesis, is that people's templates contain the full set of potentially important features at all stages of training. The APP Hypothesis predicts, in contrast, that people's templates will be characterized by small subsets of features whose members vary over time. Next, consider an experiment in which many features are equally diagnostic of a stimulus category. A global pooling hypothesis would claim that people's templates will contain all of these features. In contrast, the APP

Hypothesis predicts that these templates will contain only a small subset of features. If attention and learning are linked, then the subset will contain those features that are most salient on the basis of endogenous and exogenous selective attention.

At a neurophysiological level, recent data by Law and Gold (2008) are consistent with basic predictions of the APP Hypothesis. These authors trained monkeys to perform a visual motion-direction discrimination task. They found that performance improvements were accompanied by changes in neural responses in cortical area LIP, an area associated with decision making, but not in neural responses in area MT, an area associated with sensory representation. Moreover, during the course of training, the responses of the most sensitive neurons in MT became more strongly correlated with a monkey's decisions. These data support the APP Hypothesis because they are consistent with the idea that performance improvements are due to a more selective read-out of the most sensitive sensory signals.

The APP Hypothesis motivates additional experiments evaluating finer details of the hypothesis, such as experiments in which image regions or features to which an observer attends are manipulated. Consider, for example, a monkey discriminating two very close directions of visual motion, such as  $0^\circ$  and  $3^\circ$ . In addition, consider the subset of MT neurons whose activities are informative for this task, such as neurons that are reliably more active with  $0^\circ$  stimuli than with  $3^\circ$  stimuli. In one experimental condition, these neurons could be electrically stimulated when the monkey views  $0^\circ$  stimuli, thereby causing these neurons to become highly active. Would this manipulation make it easier for the monkey to identify these highly informative neurons? If the APP Hypothesis uses competitive learning implemented via a Hebbian mechanism, then the hypothesis predicts that the answer to this question is yes, and that the manipulation will cause the monkey's decision network to use the activities of these neurons in future decisions, thereby leading to better performance and faster learning. In another condition, area MT neurons selective for relatively uninformative motion directions could be stimulated. Would the monkey's decision network then use the activities of these neurons? Again, the hypothesis predicts that the answer is yes. In this case, the manipulation would lead to worse performance and slower learning.

Taken as a whole, the APP Hypothesis is an early step toward a neural theory of perceptual learning. In its current state, it leaves open many questions regarding, for example, its neural implementation, the relationships between learning, attention, and decision making, and its computational properties. Perhaps its greatest strength is that it provides a single, coherent framework for addressing many issues in the perceptual sciences. We hope that psychophysicists, neurophysiologists, and computational neuroscientists will all contribute to its future development.

## Acknowledgments

We thank S. Newman and M. Sacra for their help in running the experiment, and M. Chhabra, G. DeAngelis, I. Fine, A. Pouget, and R. Yakushijin for helpful discussions. We also thank two anonymous reviewers for their comments on an earlier version of this manuscript, and A. Petrov for making available on his web site the Matlab code for his representational front end. This work was supported by AFOSR Research Grant FA9550-06-1-0492 and NSF Research Grant DRL-0817250.

Commercial relationships: none.

Corresponding author: Robert A. Jacobs.

Email: robbie@bcs.rochester.edu.

Address: Center for Visual Science, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA.

## Footnotes

<sup>1</sup>There exist in the literature at least two explanations at the level of neural processing for visual learning, namely that learning results from changes in the tunings of neurons' sensitivity functions (tuning functions might shift, broaden, or sharpen) and that learning results from changes in the weights assigned to the responses of neurons contributing to a psychological response. These two explanations are compatible because reweightings of neural responses will necessarily result in changes to the tuning functions of all mechanisms (including decision mechanisms) subsequent to the reweighting.

<sup>2</sup>Our reasons for using the term "scaling" can be illustrated as follows. As discussed below, a Gaussian mixture model can be converted to an equivalent logistic regressor that maps a model's inputs to the probability that the inputs represent the image of a left-closer surface. The regressor corresponding to the Anisotropic Observer multiplies input  $i$  by  $1/\sigma_i^2$ , where  $\sigma_i^2$  is input  $i$ 's average class-conditional variance. That is, the regressor scales each input dimension by the variance along that dimension. The regressor corresponding to the Isotropic Observer multiplies all inputs by the same value, meaning that this regressor uses a uniform scaling in all input dimensions.

<sup>3</sup>Recall that a subject was trained to distinguish two surface slants ( $\tau$  and  $-\tau$ ). In addition, recall that the model neurons are tuned to specific spatial locations, frequencies, and orientations. Depending on the slant values used when training a subject, the receptive field properties of a neuron may make its responses highly diagnostic of whether a left-closer or right-closer surface was displayed on a trial, or may make it only moderately or mildly diagnostic.

## 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](#)]
- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, Optics and Image Science*, 2, 284–299. [[PubMed](#)]
- 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., 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](#)]
- Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *Journal of Physiology (London)*, 136, 469–488. [[PubMed](#)] [[Article](#)]
- Barlow, H. B. (1995). The neuron doctrine in perception. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 415–435). Cambridge, MA: MIT Press.
- 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](#)]
- Bellman, R. E. (1957). *Dynamic programming*. Princeton, NJ: Princeton University Press.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Blake, A., Bülthoff, H. H., & Sheinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research*, 33, 1723–1737. [[PubMed](#)]
- Bradley, A., Skottun, B. C., Ohzawa, I., Sclar, G., & Freeman, R. D. (1987). Visual orientation and spatial frequency discrimination: A comparison of single neurons and behavior. *Journal of Neurophysiology*, 57, 755–772. [[PubMed](#)]
- Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, 13, 87–100. [[PubMed](#)]
- Chubb, C., & Landy, M. S. (1991). Orthogonal distribution analysis: A new approach to the study of

- texture perception. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 291–301). Cambridge, MA: MIT Press.
- Chubb, C., & Talevich, J. (2002). Attentional control of texture orientation judgments. *Vision Research*, *42*, 311–330. [PubMed]
- Cumming, B. G., Johnston, E. B., & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, *33*, 827–838. [PubMed]
- Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*, 198–216. [PubMed]
- Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*, 1218–1223. [PubMed]
- Dosher, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 13988–13993. [PubMed] [Article]
- Dosher, B. A., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision Research*, *39*, 3197–3221. [PubMed]
- Eckstein, M. P., Abbey, C. K., Pham, B. T., & Shimozaki, S. S. (2004). Perceptual learning through optimization of attentional weighting: Human versus optimal Bayesian learner. *Journal of Vision*, *4*(12):3, 1006–1019, <http://journalofvision.org/4/12/3/>, doi:10.1167/4.12.3. [PubMed] [Article]
- Fahle, M. (2004). Perceptual learning: A case for early selection. *Journal of Vision*, *4*(10):4, 879–890, <http://journalofvision.org/4/10/4/>, doi:10.1167/4.10.4. [PubMed] [Article]
- Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, *35*, 3003–3013. [PubMed]
- Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*, 43–44. [PubMed]
- Földiák, P. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, *64*, 165–170. [PubMed]
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*, 267–314. [PubMed]
- Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning; differentiation or enrichment? *Psychological Review*, *62*, 32–41. [PubMed]
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, *402*, 176–178. [PubMed]
- Gold, J., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, *28*, 167–207.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*, 1157–1182.
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181–197. [PubMed]
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804. [PubMed] [Article]
- Jazayeri, M., & Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, *446*, 912–915. [PubMed]
- Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 4966–4970. [PubMed] [Article]
- Knill, D. C. (1998). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, *38*, 2635–2656. [PubMed]
- Law, C.-T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience*, *11*, 505–513. [PubMed] [Article]
- Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*, 493–497. [PubMed] [Article]
- 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]
- Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research*, *40*, 97–109. [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]

- MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.
- 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. [[PubMed](#)] [[Article](#)]
- Nadler, J. W., & DeAngelis, G. C. (2005). Precision pooling predicts primate perceptual performance. *Nature Neuroscience*, 8, 12–13. [[PubMed](#)]
- Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, 249, 1037–1041. [[PubMed](#)]
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607–609. [[PubMed](#)]
- Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annual Review of Neuroscience*, 21, 227–277. [[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](#)]
- Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review Neuroscience*, 26, 381–410. [[PubMed](#)]
- Purushothaman, G., & Bradley, D. C. (2005). Neural population code for fine perceptual decisions in area MT. *Nature Neuroscience*, 8, 99–106. [[PubMed](#)]
- Purves, D., & Lichtman, J. W. (1985). *Principles of neural development*. Sunderland, MA: Sinauer Associates.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Rubin, N., Nakayama, K., & Shapley, R. (2002). The role of insight in perceptual learning: Evidence from illusory contour perception. In M. Fahle & T. Poggio (Eds.), *Perceptual learning* (pp. 235–251). Cambridge, MA: MIT Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: What is learned? *Vision Research*, 35, 519–527. [[PubMed](#)]
- Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practicing orientation identification improves orientation coding in V1 neurons. *Nature*, 412, 549–553.
- Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience*, 16, 1486–1510. [[PubMed](#)] [[Article](#)]
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1193–1216. [[PubMed](#)]
- Tanner, W. P., & Birdsall, T. G. (1958). Definitions of  $d'$  and  $\eta$  as psychophysical measures. *Journal of the Acoustical Society of America*, 30, 922–928.
- Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *WESCON Convention Report*, 4, 96–104.