Comparing perceptual learning across tasks: A review

Ione Fine   Department of Psychology,
University of California, La Jolla, CA, USA

Robert A. Jacobs   Center for Visual Science,
University of Rochester, Rochester, NY, USA

We compared perceptual learning in 16 psychophysical studies, ranging from low-level spatial frequency and orientation discrimination tasks to high-level object and face-recognition tasks. All studies examined learning over at least four sessions and were carried out foveally or using free fixation. Comparison of learning effects across this wide range of tasks demonstrates that the amount of learning varies widely between different tasks. A variety of factors seems to affect learning, including the number of perceptual dimensions relevant to the task, external noise, familiarity, and task complexity.

Keywords: plasticity, training, pattern discrimination, object recognition, detection, discrimination, learning

Introduction

Psychophysical and neurophysiological evidence has made it increasingly obvious that the adult visual system is plastic at almost all stages of processing, from the photoreceptors (Smallman, MacLeod, & Doyle, 2001) to extrastriate areas concerned with object recognition (Kobatake, Wang, & Tanaka, 1998; Zohary, Celebrini, Britten, & Newsome, 1994). Here we examine the effect of task complexity in 16 tasks, ranging from studies of simple orientation judgments to studies of object recognition. We chose studies that were homogenous in as many methodological details as possible, and therefore only included a restricted subset of the growing number of studies on perceptual learning. We used seven main criteria for including studies. First, we limited our review to studies that examined relatively long-term learning processes by requiring that training was carried out for at least four sessions, with training sessions lasting at least 30 minutes, and only one session carried out each day. Although remarkably specific (and long lasting) learning effects have been found to take place within an hour or two of training (e.g., Fiorentini & Berardi, 1980, 1981; Shiu & Pashler, 1992; Fahle, Edelman, & Poggio, 1995; Liu & Vaina, 1998), we chose to focus on slow learning processes that take place over a number of sessions. Because of possible fatigue effects (e.g., Beard, Levi, & Reich, 1995) and the role of sleep in consolidating learning (Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994), we excluded studies where significantly more than an hour of training was carried out per day (e.g., Vogels & Orban, 1985). Second, tasks were carried out foveally or with free fixation. Data allowing comparisons between learning in the periphery and the fovea have been obtained only for low-level tasks (e.g., Johnson & Leibowitz, 1979; Fendick & Westheimer, 1983; Bennett & Westheimer, 1991; Westheimer, 2001). Because the current state of the literature did not provide enough data

to determine how learning interacts with task and eccentricity, we excluded studies carried out using stimuli that extended into the periphery (e.g., Beard et al., 1995; Westheimer, 2001; Ahissar & Hochstein, 1996). Third, we only included studies where improvements did not seem to be limited by ceiling effects (i.e., performance did not exceed 95% correct by the end of training). Fourth, we excluded studies where observers were given any significant pretraining. Fifth, we only included studies where error feedback was given after each trial because some studies show stronger learning effects when feedback is given after every trial than when no feedback is given (Herzog & Fahle, 1997; Shiu & Pashler, 1992). Sixth, tasks had to be purely perceptual. For example, in the Gauthier, Williams, Tarr, and Tanaka (1998) study on object recognition, the task involved learning arbitrary names (e.g., "vali" and "pimo") for "greebles" and parts of "greebles" (e.g., "boges" and "dunth"), and, therefore, involved a substantial semantic memory component. Finally, we only included studies where percent correct, *d'*, or thresholds were used as a performance measure. Studies using reaction time (e.g., Vidyasadar & Stuart, 1993) were excluded, as were studies using a combination of percent correct and reaction time (e.g., Gauthier et al., 1998), because encouraging subjects to respond as quickly as possible might result in a speed-accuracy trade-off.

In addition, where possible, we chose studies that used at least three observers because the size of training effects is notoriously subject to individual differences. In a few cases, we have included studies with very similar stimuli and training procedures that were carried out in different laboratories. Repetitions of training procedures that were carried out within the same laboratory have been excluded.

Despite these restrictions, the studies we included still varied significantly in their methodology. Training sessions could last anywhere between 30 and 60 minutes. In some studies, subjects were trained till asymptote, whereas in

other studies, they were trained for a fixed number of sessions. Some studies used fixed stimuli that did not vary across training sessions, whereas in other studies, stimulus intensity depended on the performance of the observer. Some studies used only naïve observers, whereas others included experienced psychophysical observers (occasionally the authors). A wide variety of tasks were used, including same-different, yes-no, 2-alternative forced choice (AltFC), 4AltFC, and match to sample.

In all studies, we converted performance into measures of *d'* before and after learning (see the attached source code for further details). Signal detection theory (SDT) has been used to interpret subjects' performances in a wide variety of perceptual and cognitive tasks, and *d'* can be calculated for a variety of measurement techniques (e.g., percent correct and threshold) and psychophysical procedures (e.g., yes-no or forced choice) as described in Green and Swets (1966) and Macmillan and Creelman (1991). Moreover, *d'* tends to be robust to violations of its assumptions (in some circumstances this may be due to the central limit theorem and SDTs frequent use of normal distributions). For these reasons, we chose *d'* as a reasonable candidate for a common metric that could be used to compare learning across studies. We used a learning index, *L*, as our measure of improvements in performance with practice, $L_s = d'_s/d'_{s=1}$, where *s* is the session number. Similar indices have been used to measure attentional effects in neurophysiology and fMRI studies (Treue & Maunsell, 1996; Gandhi, Heeger, & Boynton, 1999). The larger the learning index, the greater the amount of learning: a learning index remaining near 1 implies that observers showed no improvement in performance with practice. Learning is generally modeled with an exponential function, because at some point performance necessarily asymptotes. However, in many studies, performance never approached asymptote, and over the first four sessions, we found that learning, measured using *d'* was better fit by a linear rather than by an exponential function. We estimated the slope of learning ($slope_L$), by fitting a line to the learning indices L_s for s={1, 2, 3, 4}. No learning would result in a slope of 0, whereas *d'* doubling across each session would result in a slope of 1. We based our estimation of the slope on data from the first four sessions for two reasons. First, all of the studies included took place over at least 4 days, and second, in a few studies, observers' behavior seemed to begin to be asymptote by the fifth session. Unfortunately, the data available to us made it impossible to reliably compare asymptotes between studies (see Figure 3).

Though some of the studies in this paper may not have strictly complied with the assumptions made by signal detection theory, our estimates of learning were remarkably robust to deviations in the assumptions that we made. For example, simulations showed that our estimates of learning were very robust to variation in our estimates of the relative standard deviations of signal and noise. Simulations also showed that, within reasonable

limits, our estimates of learning were reasonably robust to deviations from the assumption that observers always used the best possible criterion. When calculating changes in *d'* with practice, we chose stimulus intensities well within the mid range of the psychometric curves describing performance before and after practice. When converting threshold measures to *d'*, we chose a stimulus intensity where *d'* was between 0.5 and 1 at the beginning of training (corresponding to a stimulus intensity resulting in performance between 59.9%-68.7% correct in a yes-no task). Figure 1A and 1B show hypothetical curves for percent correct and *d'* as a function of stimulus intensity before (solid line) and after (dashed line) training in a yes-no task. The red arrows indicate changes in percent correct and *d'* for a stimulus intensity corresponding to *d'=0.5* at the beginning of training; at the end of training, *d'* was 2.3, corresponding to *L=4.7*. The blue arrows indicate changes in percent correct and *d'* for a stimulus intensity corresponding to *d'=1* at the beginning of training; by the end of training, *d'* was 4.1, corresponding to *L=4.1*. Conveniently, simulations showed that provided thresholds and slopes fell within reasonable limits, our estimation of the learning index was fairly robust to the choice of the intensity value for which we calculated changes in *d'*, especially for smaller values of *L*, for example, when *L≈1* estimates vary by 0.5% to 1% depending on whether *d'=0.5* or *d'=1* was chosen as a starting point. For *L≈2*, estimates vary by about 3% to 6%, for *L≈4*, estimates vary by about 15%, and for *L≈6*, estimates vary by about 20%.
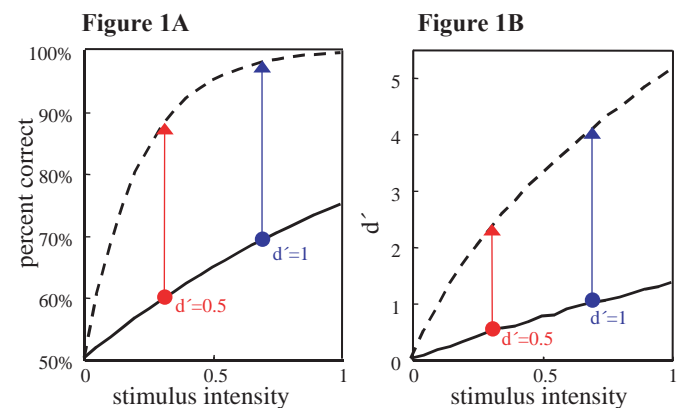

Figure 1A. Hypothetical curves showing percent correct as a function of stimulus intensity before (solid line) and after (dashed line) training for a yes-no task. B. *d'* as a function of stimulus intensity (arbitrary units) before and after training. Red arrows indicate changes in percent correct and *d'* for a stimulus intensity corresponding to *d'=0.5* at the beginning of training, and the blue arrows indicate changes in percent correct and *d'* for a stimulus intensity corresponding to *d'=1* at the beginning of training.

(1) Matthews, Liu, Geesaman &
Qian, 1999

(2) Johnson and Leibowitz (1979)

(3a) Ball & Sekuler, 1987
(3b) Matthews and Welch, 1997

(4a) Matthews & Welch, 1997
(4b) Matthews, Liu, Geesaman
& Qian, 1999

(5) Fine & Jacobs, 2000

(6) Furmanski & Engel, 2000

(7a) Ball & Sekuler, 1987
(7b) Matthews and Welch, 1997

(8a) Matthews & Welch, 1997
(8b) Matthews, Liu, Geesaman
& Qian, 1999

(9) Fine & Jacobs, 2000

(10) Matthews & Welch, 1997

(11) Herzog & Fahle, 1997

(12) Gold, Bennett &
Sekuler, 1999

(13) Gold, Bennett &
Sekuler, 1999

(14) Gold, Bennett
& Sekuler, 1999

(15) Sigman and Gilbert, 2000
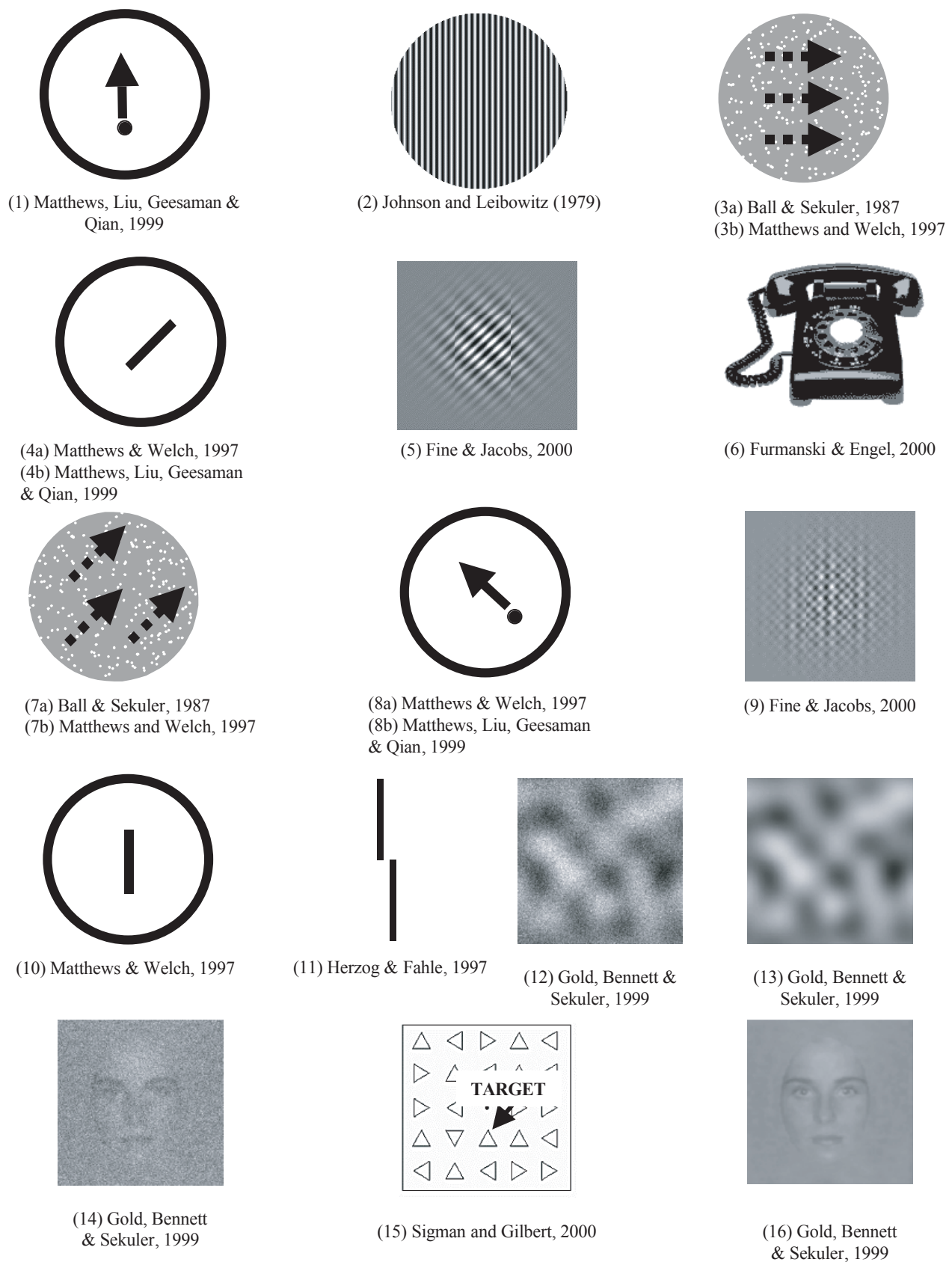
(16) Gold, Bennett
& Sekuler, 1999

Figure 2. Examples of the stimuli used in the 16 tasks described above.
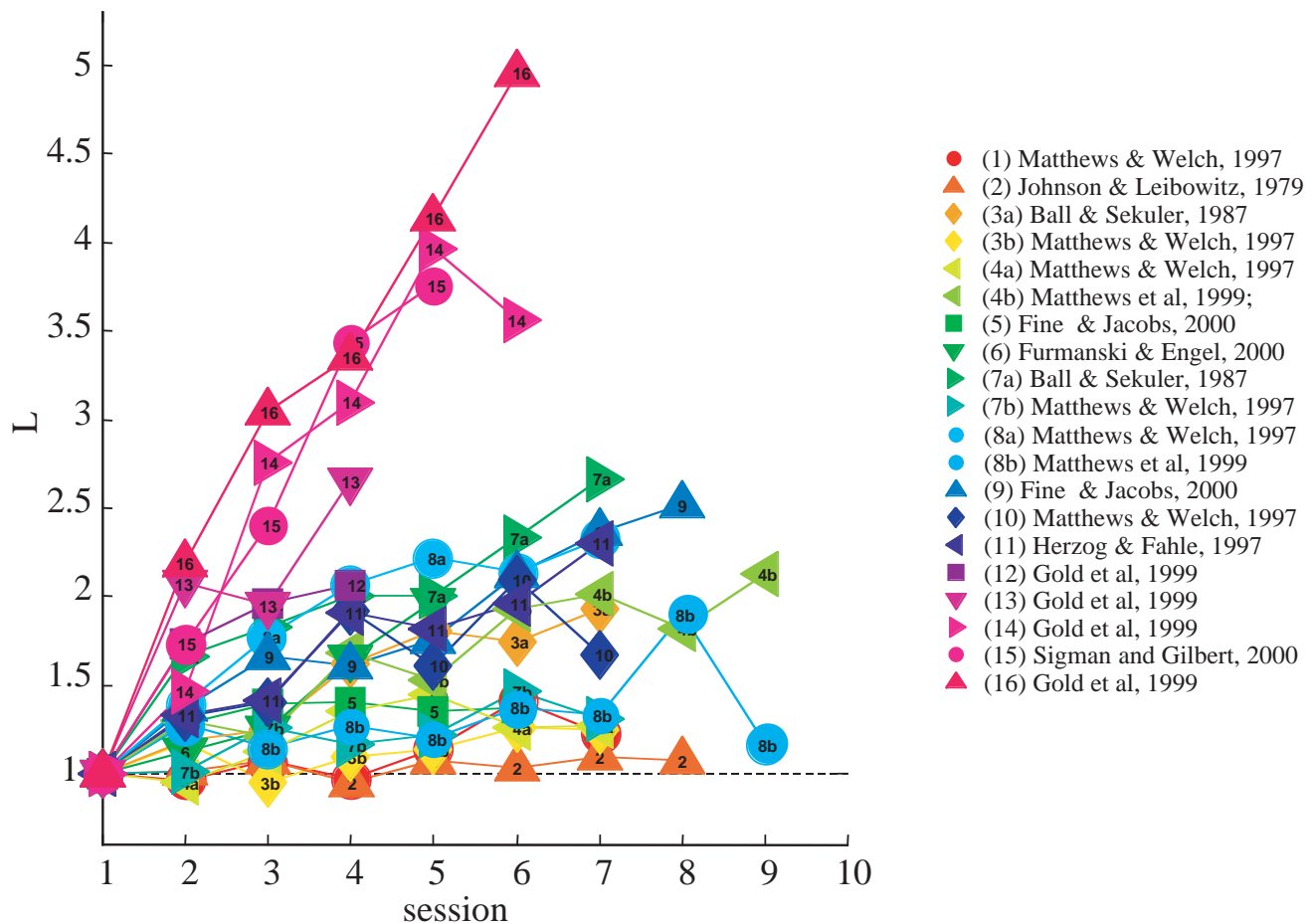
Figure 3. Learning (L) as a function of session for each of the 16 tasks.

List of Studies

Tasks are listed below and in Figure 2, in ascending order, according to the estimated slope of learning ($slope_L$). Subjects showed the least learning in the tasks described in the beginning of this section, and the most learning in the tasks described at the end of this section. Figure 3 shows learning as a function of session for each study. Where we included more then one study using very similar stimuli and procedures, we have listed them according to the mean slope of learning averaged across the different studies.

1. Cardinal direction of motion discrimination for a single dot

Matthews and Welch (1997) trained observers to discriminate differences in the direction of motion for a single moving dot moving within a 10-degree aperture. The direction of motion was 0° or 90° and the dot traveled at 2, 10, or 16 degrees/s. Observers were presented with a moving dot stimulus in each of two temporal intervals, and were asked to indicate whether the direction of motion in the second interval was rotated clockwise or counterclockwise compared to the first.

Performance is averaged across five observers. Observers showed almost no learning; $slope_L$ was 0.001.

2. Resolution limit for gratings

Johnson and Leibowitz (1979) measured observers' resolution limits for sinusoidal gratings windowed within a 2-degree circular aperture using a yes-no forced choice procedure. Performance is averaged across four observers. Observers showed almost no learning; $slope_L$ was 0.002.

3. Cardinal direction of motion discrimination for a field of dots

(a) Ball and Sekuler (1982, 1987) trained observers to discriminate small changes in the direction of motion of a field of spatially random dots moving with 100% motion coherence. Observers were presented with stimuli in two temporal intervals, and had to report whether the direction of motion in the two intervals was the same or different. The dots moved in one of four cardinal directions (centered on 0°, 90°, 270°, and 180°). The direction of motion difference between the two intervals was 3°, and was randomly selected to be either clockwise or counterclockwise. Performance is averaged across 8 observers; $slope_L$ = 0.183.

(b) Matthews and Welch (1997) carried out a very similar study in which observers were trained to discriminate differences in the direction of motion for a field of random dots moving within a 10-degree aperture. The direction of motion was 0° or 90° and the dot traveled at 2, 10, or 16 degrees/s (the data presented here are averaged over all three speeds). Observers were presented with a moving field of dots in each of two temporal intervals, and were asked to indicate whether the direction of motion in the second interval was rotated clockwise or counterclockwise compared to the first. Performance is averaged across six observers. Observers showed only a small amount of learning; the slope of learning, $slope_L$, was 0.0261. $Slope_L$, averaged across both studies (3a, 3b), was 0.1046.

4. Oblique orientation discrimination

(a) Matthews and Welch (1997) trained observers to discriminate orientation differences between two single-line stimuli. Each line stimulus was 1, 5, or 8 degrees long and 5 min wide, and had an orientation of 45° or 135°. Observers were presented with a line stimulus in each of two temporal intervals, and were asked to indicate whether the second stimulus was rotated clockwise or counterclockwise compared to the first. Performance is averaged across five observers; $slope_L$= 0.0903.

(b) Similarly, Matthews, Liu, Geesaman, and Qian (1999) trained observers to discriminate orientation differences between two single-line stimuli. Each line stimulus was 2 degrees long and 5 min wide, and had an orientation of 45° or 135°. Observers were presented with a line stimulus in each of two temporal intervals, and were asked to indicate whether the second stimulus was rotated clockwise or counterclockwise compared to the first. Performance is averaged across five observers; $slope_L$= 0.1994. $Slope_L$ averaged across both studies (4a, 4b), was 0.1449.

5. Spatial frequency discrimination for a simple plaid

Fine and Jacobs (2000) asked observers to discriminate changes in spatial frequency within a simple plaid pattern using a 4AltFC task. The plaid contained two orthogonal gratings with spatial frequencies near 3 and 9 cycles/degree (cpd) and respective contrasts of 3.2% and 11%. Observers were asked to discriminate which of four temporal intervals contained a slight shift in spatial frequency within both gratings in the plaid. Phases were randomized in each interval. Performance, averaged across three observers, showed a small amount of learning; $slope_L$ = 0.1631,

6. Familiar object identification

Furmanski and Engel (2000) asked observers to identify common objects. Observers were asked to name gray-scale images of briefly presented common objects (e.g., clock, brush, and stapler). Each observer was trained

on 20 objects. Each session began with a series of 2-s displays in which each of the 20 objects was presented along with its name. Stimuli were then briefly presented and observers were asked to name the object. Performance shown here is averaged across four observers. Three replications of this, or a very similar training procedure, resulted in very similar learning effects. Observers showed a small amount of learning; $slope_L$ = 0.1836.

7. Oblique direction of motion discrimination for a field of dots

(a) Ball and Sekuler (1982, 1987) trained observers to discriminate small changes in the direction of motion of a field of spatially random dots moving with 100% motion coherence. Observers were presented with stimuli in two temporal intervals, and had to report whether the direction of motion in the two intervals was the same or different. The dots moved in one of four oblique directions (centered on 45°, 135°, 225°, and 315°). The direction of motion difference between the two intervals was 3 degrees, and was randomly selected to be either clockwise or counterclockwise. Performance is averaged across eight observers; $slope_L$ = 0.381.

(b) Similarly, Matthews and Welch (1997) carried out a study in which observers were trained to discriminate differences in the direction of motion for a field of random dots moving within a 10-degree aperture. The direction of motion was 45° or 135° and the field of dots traveled at 2, 10, or 16 degrees/s (the data presented here are averaged over all three speeds). Observers were presented with a moving field of dots in each of two temporal intervals, and were asked to indicate whether the direction of motion in the second interval was rotated clockwise or counterclockwise compared to the first. Performance is averaged across six observers. Observers showed only a small amount of learning; the slope of learning, $slope_L$ was 0.0727. $Slope_L$, averaged across both studies (7a, 7b) was 0.2269.

8. Oblique direction of motion discrimination for a single dot

(a) Matthews and Welch (1997) trained observers to discriminate differences in the direction of motion for a single moving dot moving within a 10-degree aperture. The direction of motion was 45° or 135° and the dot traveled at 2, 10, or 16 degrees/s (the data presented here are averaged over all three speeds). Observers were presented with a moving dot stimulus in each of two temporal intervals, and were asked to indicate whether the direction of motion in the second interval was rotated clockwise or counterclockwise compared to the first. Performance is averaged across six observers. Observers showed only a small amount of learning; the slope of learning, $slope_L$ was 0.3676.

(b) In a very similar study, Matthews et al. (1999) trained observers to discriminate differences in the

direction of motion for a single moving dot moving within a 10-degree aperture at 10 degrees/s. The direction of motion was 45° or 135°. Observers were presented with a moving dot stimulus in each of two temporal intervals, and were asked to indicate whether the direction of motion in the second interval was rotated clockwise or counterclockwise compared to the first. Performance is averaged across five observers. Observers showed only a small amount of learning; the slope of learning, $slope_L$, was 0.0979. $Slope_L$ averaged across both studies (8a, 8b) was 0.2327.

9. Spatial frequency discrimination for a complex plaid

Fine and Jacobs (2000) asked observers to discriminate changes in spatial frequency within a complex plaid pattern using a 4AltFC task. The "wicker" texture contained two orthogonal signal gratings masked by four noise gratings. One signal grating was centered on 3 cpd, had an orientation of -45°, and a contrast of 1.5% to 12.8%. The other signal grating was centered on 9 cpd, had an orientation of 45°, and a contrast of 5.5% to 44%. The first noise component had a frequency of 9 cpd, an orientation of -45 degrees, and a contrast of 11.2%. The second noise grating had a frequency of 3 cpd, an orientation of 45 degrees, and a contrast of 3.2%. The third noise grating had a frequency of 4.3 cpd, an orientation of 0 degree, and a contrast of 7.1%. The fourth noise grating had a frequency of 6.2 cpd, an orientation of 9 degrees, and a contrast of 7.1%. Phases were randomized in each presentation interval. Observers were asked to discriminate in which of four temporal intervals both signal gratings in the plaid shifted slightly in spatial frequency. Observers showed more improvement than when asked to discriminate small changes in spatial frequency within simple plaids (study 5), suggesting that integrating information across a wide range of spatial frequencies and orientations is a relatively plastic process. Observers showed relatively large improvements in performance as they learned to base their responses on the spatial frequencies and orientations that are relevant for the task. $Slope_L$ averaged across five observers was 0.2517.

10. Cardinal orientation discrimination

Matthews and Welch (1997) trained observers to discriminate orientation differences between two single-line stimuli. Each line stimulus was 1, 5, or 8 degrees long and 5 min wide and had an orientation of 0° or 90°. Observers were presented with a line stimulus in each of two temporal intervals, and were asked to indicate whether the second stimulus was rotated clockwise or counterclockwise compared to the first. Performance is averaged across five observers; $slope_L$= 0.2785.

11. Vernier offset discrimination

Many learning studies have examined performance discriminating Vernier offsets (e.g., McKee & Westheimer, 1978; Fahle & Edelman, 1993; Fahle et al., 1995) with the belief that this is a task mediated by fairly low-level visual mechanisms. Herzog and Fahle (1997) used two straight lines (10 × 2 arc min) that were slightly displaced relative to each other, and trained subjects to discriminate the direction of the offset. The presentation time was 150 msec. Half the observers performed the task using horizontal lines as stimuli, the other half using vertical lines. Performance is averaged across 10 observers and both orientations; $slope_L$= 0.290.

12. Band-pass noise identification with high-contrast noise

Gold, Bennett, and Sekuler (1999) examined the ability of observers to discriminate between 10 band-pass Gaussian filtered noise textures. The textures were Gaussian noise fields (5.25 × 5.25 degrees) filtered by a 2 to 4 cycle per image rectangular frequency filter. High-contrast external two-dimensional noise, with a spectral density of 25.55×10^{-6} deg^2, was added to each noise texture to make discrimination more difficult. Each texture, with added noise, was displayed for 500 msec. Observers identified each texture as one of a set of noise-free versions of each texture. Performance is averaged across two observers; $slope_L$ = 0.4195.

13. Band-pass noise identification with low-contrast noise

Gold et al. (1999) examined the ability of observers to discriminate between 10 band-pass Gaussian filtered noise textures. The textures were Gaussian noise fields (5.25 × 5.25 degrees) filtered by a 2 to 4 cycle per image rectangular frequency filter. Low-contrast external two-dimensional noise, with a spectral density of 0.04×10^{-6} deg^2, was added to each noise texture to make discrimination more difficult. Each texture, with added noise, was displayed for 500 msec. Observers identified each texture as one of a set of noise-free versions of each texture. Performance is averaged across two observers; $slope_L$ = 0.5666.

14. Novel face discrimination with high-contrast noise

Gold et al. (1999) examined the ability of observers to discriminate between 10 faces. High-contrast external two-dimensional noise, with a spectral density of 25.55×10^{-6} deg^2, was added to each face to make discrimination more difficult. Each face, with added high-contrast noise, was displayed for 500 msec. Observers matched the stimulus face to a set of noise-free versions of every face. Performance is averaged across two observers; $slope_L$ = 0.7350.

15. Simple shape search

Sigman and Gilbert (2000) asked observers to report whether a randomly positioned triangle was present within a display of 24 distracters. Observers were trained with the target triangle at a particular cardinal orientation, with the distracter triangles oriented along the other three cardinal axes. The sides of the triangles were 27 min in length and their centers were separated by 54 min. The stimulus array subtended 4.2×4.2 degrees of visual angle, and a small fixation spot of 1 arc min radius was positioned in its center. The stimulus array was presented for 300 msec on every trial. Performance is averaged across four observers; s$lope_L$ = 0.7771.

16. Novel face discrimination with low-contrast noise

Gold et al. (1999) examined the ability of observers to discriminate between 10 faces. Low-contrast external two-dimensional noise, with a spectral density of 0.04×10^{-6} deg^2, was added to each face to make discrimination more difficult. Each face, with added low-contrast noise, was displayed for 500 msec. Observers matched the stimulus face to a set of noise-free versions of every face. Performance is averaged across two observers; $slope_L$ = 0.8815.

Discussion

As can be seen from Figure 3, the amount of learning varies widely between different tasks. Some tasks (e.g., cardinal direction discrimination for a single dot and resolution limits) show no or almost no improvement with practice, whereas in other tasks (e.g., novel face discrimination and shape search) d' improved by more than a factor of three over four sessions of training.

It is still not clear what sort of neuronal changes underlie these improvements in performance found with practice. One suggestion is that perceptual learning might be mediated by changes in the tuning of the sensitivity functions of the relevant neurons: neural tuning functions might shift, sharpen, or broaden with practice depending on the stimulus and the task. Alternatively, it has been suggested that learning might be a consequence of selective reweighting of the neurons that contribute to the psychophysical response, so that the neurons best tuned for optimal performance are given more weight (e.g., Saarinen & Levi, 1995). We believe that these two explanations are consistent with each other, because selective reweighting of neurons will necessarily result in changes in the tuning functions of all mechanisms (including decision mechanisms) subsequent to the reweighting. It seems likely that this reweighting or retuning as a function of practice may not result in permanent changes in the tuning properties of neurons, but may instead be context dependent. The lack of transfer across stimuli and tasks found psychophysically

(e.g., Beard et al., 1995), as well as the context-dependent learning effects found by Crist, Li,, and Gilbert (2001), suggests that even at very early stages of processing, reweighting may be task specific and mediated by higher-level cognitive feedback and attention.

Obviously the neuronal changes underlying performance improvements may well differ substantially depending on the task. For example, the process of reweighting of inputs (and the consequent shifts in tuning) may take place sequentially throughout the visual system. Consistent with this, observers often seemed to show more learning for the stimuli that intuitively might be considered more complex (Green & Swets, 1966). Figure 2 shows the stimuli from the different tasks, ranked in order of the learning slope, with the tasks that showed least learning at the top. As can be seen from Figure 2, tasks involving relatively simple stimuli (plaids, bars, moving dots) and judgments along a single perceptual dimension, such as a spatial frequency, orientation, or direction of motion, tended to show only small amounts of learning.

We classified tasks as low level if they involved a judgment along a "basic" perceptual dimension, such as a single spatial frequency, orientation, direction of motion, or position. In none of the tasks described above were the stimuli corrupted by external noise (external noise paradigms have tended to be carried out in the periphery where learning effects are larger). Eleven tasks were classified as low level: resolution limit thresholds (study 2), direction of motion discrimination for a single dot (studies 1, 8a & 8b), direction of motion discrimination for a field of dots (studies 3a, 3b, 7a & 7b), orientation discrimination (studies 4a, 4b & 10), and Vernier offset discrimination (study 11). Performance on low-level tasks showed fairly limited improvement with practice; after four sessions, the slope of the learning index averaged across these tasks (treating different studies using the same task as separate studies) was $slope_L$ = 0.1658. This limited learning is a little surprising given the growing literature on low-level plasticity. However, many low-level learning studies finding strong learning effects have examined improvements in performance within one or two sessions (e.g., Fiorentini & Berardi, 1980, 1981; Shiu & Pashler, 1992; Fahle et al., 1995) or have examined learning in the periphery (Crist, Kapadia, Westheimer, & Gilbert, 1997; Dosher & Lu, 1998, 1999; Mayer, 1983) where, as described above, learning effects seem to be larger (Fendick & Westheimer, 1983; Bennett & Westheimer, 1991; Westheimer, 2001), at least for low-level tasks.

Often learning for these low-level stimuli is very specific for orientation, spatial position, size, and, occasionally, eye of origin. It has, therefore, been argued that learning must be taking place in neurons, situated early in processing, that are selective for these properties. However, it is possible that neurons normally unselective for properties like orientation and spatial position might

become more selective with training (Mollon & Danilova, 1996). Unfortunately, the neurophysiological evidence for changes as a function of practice at early stages of visual processing is still fairly weak. Although Schoups, Vogels, Qian, and Orban (2001) have evidence for small changes in the slope of neural tuning within V1 as a function of practice using an orientation discrimination task in the periphery, other studies have not found evidence for significant changes in population-tuning properties using a bisection task (Crist et al., 2001) and an orientation-discrimination task very similar to that of Schoups et al. (Ghose, Yang, & Maunsell, 2002). However, Crist et al. did find context-dependent surround interactions within V1 after training, suggesting that practice modifies task-dependent feedback from higher visual areas.

We found that the five tasks using stimuli that contained external noise (studies 9, 12, 13, 14, and 16) showed, on the whole, more learning than low-level tasks, with the slope of the learning index averaged across studies containing external noise being $slope_L = 0.5709$. This observation is not particularly surprising because several studies have found greater learning effects when external noise is added to a stimulus (Dorais & Sagi, 1997; Dosher & Lu, 1998, 1999; Saarinen & Levi, 1995; Gold et al., 1999; though curiously in the Gold et al. studies, more learning was demonstrated in low- than in high-noise conditions). Reweighting or retuning of neurons would help to exclude external noise (by reducing the weighting of those neurons for which tuning does not match the stimulus well, or for which responses are particularly sensitive to external noise) as well as to reduce internal noise by excluding neurons with a low signal-to-noise ratio. Improvements in performance on a variety of tasks have been shown to be due to a combination of external noise exclusion and, in some studies, to suppression of internal noise. For example, Dorais and Sagi (1997), Dosher and Lu (1998, 1999), and Saarinen and Levi (1995), using orientation discrimination, contrast detection, and Vernier acuity tasks, have found that learning for stimuli masked with external noise is consistent with external noise exclusion as a major factor in perceptual learning. Gold et al. (1999), using a similar external noise technique on face recognition, also found that a significant amount of learning seemed to be due to a reduction in external noise, with training seeming to have little effect on internal noise.

We also found that more complex tasks that required discriminations along more than one perceptual dimension showed more learning. Five tasks required discrimination between patterns containing more than one spatial frequency and orientation (studies 5, 9, 12, and 13) or a shape discrimination (study 15). The average slope of learning for these studies was $slope_L = 0.4356$ (note that three of these stimuli also included external noise). Modifying the tuning of neurons or placing more weight on the outputs of neurons best tuned for a task may be

more difficult when the useful information in a stimulus varies along multiple perceptual dimensions.

We classified tasks as high level if they involved identifying or discriminating real-world natural objects. Three tasks were categorized as high level: familiar object recognition (study 6) and novel face recognition with low- and high- contrast noise (studies 14 & 16). The average slope of learning across these three tasks was $slope_L = 0.6$. Although observers showed large amounts of learning in an unfamiliar face-identification task, they showed much less learning for familiar objects. One possibility is that previous experience of observers with the familiar objects used in the Furmanski and Engel (2000) experiment may have limited the extent of further learning within the experiment. Observers may have already had mechanisms that were optimally (or close to optimally) tuned for identification of objects at the basic level of categorization used in the study. Consequently, performance may have been limited mainly by factors such as irreducible internal noise, limiting the potential for further improvement. Tasks using familiar stimuli generally demonstrate less learning than those using less familiar stimuli: for example, Ball and Sekuler (1982, 1987; see above, studies 3a and 7a); Matthews and Welch (1997; see above, studies 3b and 7b); and others (Mayer, 1983; Vogels & Orban, 1985; McKee & Westheimer, 1978) have found more learning for oblique as opposed to cardinally oriented stimuli.

Consistent with the tendency for more complex tasks to show more learning, neuronal tuning at higher stages of processing has been shown to be highly experience dependent. This experience-dependent plasticity may help alleviate the trade-off between the need to have highly specific neurons, and biological limitations on the number of neurons that can be devoted to visual processing. Despite the fact that it would require a prohibitive number of neurons to represent every possible stimulus (because the more selective a neuron is, the smaller the number of possible stimuli it can represent), neuronal tuning in extrastriate cortex is remarkably specific. It seems that rather than representing every possible stimulus, neurons only represent a subset of possible stimuli. Especially at higher stages of processing, selectivity seems to be strongly shaped by experience, with neurons preferentially representing stimuli that have been frequently encountered, or behaviorally important in the past. Given that past experience is a good predictor of future experience, adaptability allows neurons to selectively represent an ecologically important subset of all possible stimuli. For example, neurons in inferotemporal cortex (IT) are not strongly tuned for retinotopic position, but are tuned for particular shapes: for example, neurons in macaque IT respond to particular objects and shapes, including hands and faces (Desimone, Albright, Gross, & Bruce, 1984; Logothetis, Pauls, & Poggio, 1995). These responses seem to be strongly shaped by experience with objects particular to that animal's environment. For

example, monkey face selective cells in IT show different responses to different faces, with their responses carrying identity information. The tuning of these cells seems also to be dependent on factors other than physical similarity, such as familiarity or social hierarchy (Young & Yamane, 1992; Rolls & Tovee, 1995).

Appendix A

This Appendix includes a brief summary of signal detection theory (based mainly on Green & Swets, 1966) with explanations of the assumptions and methods used in our calculations.

SDT in a Yes-No Psychophysical Procedure

In a typical yes-no psychophysical task, an observer is presented with an observation interval that contains noise (*n*) alone or contains both signal and noise (*s*). The observer responds *yes* (S) if she believes the signal was present and *no* (N) otherwise. *e* is the sensory event associated with the observation interval. *P(s)* is the a priori probability of the signal, and *P(s|e)* is the a posteriori probability that signal occurred, given the evidence *e*. Using Bayes rule,

$$P(s|e) = \frac{P(s)P(e|s)}{P(n)P(e|n) + P(s)P(e|s)}. \tag{1}$$

In such tasks, observers necessarily have a criterion (β_p) for responding S and N, based on the evidence provided by the observation interval. So for a given criterion, we can describe our subject's behavior as follows:

$$\begin{cases} \text{if: } P(s|e) \geq \beta_p \text{ , respond S} \\ \qquad \text{else: respond N} \end{cases}. \tag{2}$$

For example, in the extreme case, if an observer were rewarded for saying *yes* correctly, and was not penalized for saying *yes* incorrectly, she might choose the criterion $\beta_p = 0$, and say S on all trials, regardless of the sensory evidence (*e*).

P(S|s) is the probability of a hit: saying *yes* when the signal was present. *P(S|n)* is the probability of a false alarm: saying *yes* when only noise was present. *P(N|s)* is a miss, and *P(N|n)* is a correct rejection. A receiver-operating curve (ROC) shows how the probability of hits and false alarms change as an observer bases her responses on different criteria. As the observer lowers her criterion, the number of hits increases, but so do the number of false alarms. Because an observer only has the choice of responding yes or no, *P(S|s)+P(N|s)=1* and *P(S|n)+P(N|n)=1*. The ROC curve, therefore, also describes the number of misses and correct rejections. If signal and noise are equally likely, and the observer

chooses a criterion that maximizes the probability correct, then the probability correct is simply *p(c)=P(S|s)* or equally *p(c)=P(N|n)*.

The likelihood ratio $l_{sn}(e)$ provides a measure of the probability of evidence *e* given that the signal occurred, relative to the probability of *e* given that noise occurred:

$$l_{sn} = \frac{P(e|s)}{P(e|n)}. \tag{3}$$

Note that the likelihood ratio is independent of the a priori probability of signal and noise. The likelihood ratio is monotonically related to the a posteriori probability, provided the a priori probabilities are not zero. Because the two scales are monotonically related, criterions based on a posteriori decision rules (β_P) and the more conventionally used likelihood ratio (β) are related. For example, when signal and noise are equally probable (i.e., *P(s)=P(n)=0.5*) it can be shown that

$$l_{sn} = \frac{P(s|e)}{1 - P(s|e)}, \tag{4}$$

and a likelihood ratio criterion of β has an exact equivalent in terms of a posteriori probabilities, such that

$$\beta = \frac{\beta_p}{1 - \beta_p}. \tag{5}$$

Conveniently, in a yes-no task, the slope of the ROC curve at any point is equal to the likelihood ratio criterion that generated that point.

In many psychophysical procedures, correct decisions (hits and correct rejections) are equally rewarded, and errors (false alarms and misses) are equally penalized. In this case, the optimal decision rule is to choose a criterion that maximizes the number of hits and minimizes the number of false alarms, i.e., maximizes *P(S|s)-P(S|n)* (where noise and signal are equally likely). The best strategy is to choose S if and only if $l_{sn}(e)>=\beta$. Where false alarms were not measured, we assume in our analysis that observers weight hits and correct rejections equally and false alarms and misses equally. In all the studies we reviewed, error feedback did not distinguish between hits and correct rejections or between false alarms and misses.

SDT in a Forced-Choice Psychophysical Procedure

Most forced-choice procedures have two observation intervals, one of which contains both signal and noise (*s*) and the other of which contains noise alone (*n*). We assume that the observer's decision about which of the two intervals contains the signal is based on the likelihood ratio for each observation interval, $l_{sn}(e_i)$, *i=1, 2*, and that the two observation intervals can be treated as statistically independent.

We assume that the observer chooses the first interval if and only if the likelihood ratio associated with the first interval is greater than the likelihood ratio associated with the second interval ($l_{sn}(e_1) > l_{sn}(e_2)$). The percentage correct in a 2-alternative forced-choice task is then the area under the yes-no ROC curve,

$$P(C) = \int_0^1 \left[1 - P_k(S|n) \right] dP_k(S|s) . \tag{6}$$

The percent correct in an m-alternative forced-choice task is

$$P(C) = \int_0^1 [1 - P_k(S|n)]^{m-1} dP_k(S|s) ; \tag{7}$$

see Green and Swets (1966) for further details.

Signal and Noise Distributions

The relationships described above do not depend on the distributions of signal and noise. However, the shape of the ROC curve is heavily dependent on what assumptions are made about signal and noise. If a sensory event is thought of as being composed of many smaller, independent events, then regardless of the distribution of these underlying events, the sum of these smaller events, mapped onto a single dimension, will have a Gaussian distribution (based on the central limit theorem). Experimental evidence also suggests that the Gaussian assumption seems to hold for a wide variety of psychophysical tasks. By accepting the Gaussian assumption, signal and noise distributions can be described as

$$f(x|s) = \left(2\pi\sigma_s^2 \right)^{-1/2} \exp\left(-\frac{(x-m_s)^2}{2\sigma_s^2} \right) \tag{8}$$

and

$$f(x|n) = \left(2\pi\sigma_n^2 \right)^{-1/2} \exp\left(-\frac{(x-m_n)^2}{2\sigma_n^2} \right), \tag{9}$$

where m_s, σ_s and m_n, σ_n are the means and standard deviations of the signal and noise distributions.

In the case of a yes-no task, the ROC curve can easily be determined from these signal and noise distributions. For a given criterion k, the probability of a hit, $P(S|s)$, or a false alarm, $P(S|n)$, can be found by integrating the area under the signal or the noise distribution that falls above that criterion.

$$P(S|s) = \int_k^\infty f(x|s) dx \tag{10}$$

or

$$P(S|n) = \int_k^\infty f(x|n) dx \tag{11}$$

Because the signal and noise distributions are not directly observed, it is possible to scale the underlying variable x so that $m_n=0$, and $\sigma_n=1$, using the transformation

$$x' = \frac{x - m_n}{\sigma_n} . \tag{12}$$

Under this transformation, d', a measure of the discriminability of the signal from noise, is the difference between the means of the signal and noise distributions, divided by the standard deviation of the signal distribution,

$$d' = \frac{m_s - m_n}{\sigma_s} . \tag{13}$$

In a forced-choice task, the observer must decide which of m intervals contains the signal. One way of analyzing 2-alternative forced-choice experiments is to assume that yes-no decisions are based on the magnitude of x, whereas forced-choice decisions are based on differences in magnitude between interval 1 and interval 2, x_1-x_2. The probability of an observer responding that the signal occurred in the first interval (R_1), given that the signal occurred in the first interval (<sn>), can be expressed as,

$$P(R_1 | < sn >) = \int_k^\infty f(x_1 - x_2 | < sn >) \tag{14}$$

(see Equation 10).

Equally, the probability of an observer responding that the signal occurred in the first interval (R_1), given that the signal occurred in the second interval (<ns>), can be expressed as,

$$P(R_1 | < ns >) = \int_k^\infty f(x_1 - x_2 | < ns >) \tag{15}$$

(see Equation 11).

The resulting ROC curve is similar to the yes-no ROC curve, however $d'_{ROC} = \sqrt{2}d'$. With a few more assumptions, the ROC curve for an m-alternative forced-choice task can also be approximated (Swets, 1964).

It should be noted that the shape of the ROC is heavily dependent on the assumptions that are made about the separation between signal and noise and the relative standard deviations of the signal and noise. However, simulations showed that although d' varies if we change the relative standard deviations of signal and noise, our estimate of learning, L, remains very robust to variation in the estimate of the relative standard deviations of signal and noise (assuming that the relative standard deviations of noise and signal remain constant throughout training, which may of course, not be the case).

Percent Correct as a Function of d'
for Various Experimental Designs

Assuming that signal and noise have equal standard deviations, and that the observer maximizes the percent correct, we can calculate how percent correct is related to d' for any given task.

Threshold studies measure what stimulus intensity is necessary to achieve a fixed level of performance (usually ~75% in a 2-alternative forced-choice task). Studies measuring d' and percentage correct, on the other hand, measure performance for a fixed stimulus intensity level. Figure 1A shows two idealized psychometric functions for a yes-no task, measured before (solid) and after (dashed) training. Each value of percent correct in both psychometric functions can easily be converted into d', as described above. Figure 1B shows d' as a function of

stimulus intensity for both psychometric functions. The change in d' with practice for a particular stimulus intensity corresponds to the vertical separation between the two curves at that intensity value. As shown by the red and blue lines in Figure 1, the change in d' with practice depends on the particular stimulus intensity that is chosen. We calculated changes in d' choosing stimulus intensities well within the mid range of each psychometric curve. Where possible, we used a stimulus intensity where d' was 0.5 at the beginning of training. A d' of 0.5 corresponds to a stimulus intensity resulting in performance at 59.9% correct in a yes-no task. Conveniently, within reasonable limits, our estimate of the value of the learning index remained fairly robust to the choice of the intensity value for which we calculated changes in d' with practice.

Appendix B

The following is a MATLAB program that calculates d' from psychometric functions. Some of the routines in our simulations made use of the psychophysics toolbox (Brainard, 1997; Pelli, 1997).

```
% ExampleMain.m

% example code showing how to calculate d prime from psychometric functions
% other necessary functions are
%       FitdvPercent.m
%       dvpercent.m
%       normcdf.m

% contrast values for the psychometric function
contrast=0:.2:1;
task='YESN';

% theoretic percent correct for each contrast, before and after training
per_correct_before=[0.5000    0.5371    0.6326    0.7500    0.8542    0.9271];
per_correct_after= [0.5000    0.6326    0.8542    0.9688    0.9964    0.9998];

%initial estimate of separation
init_dprime=1;

%find dprime for each contrast
for i=1:length(contrast)
% IF YOUR MACHINE DOESN'T HAVE FMINS TRY USING FMINSEARCH %- EQUIVALENT FUNCTIONS
      if(1)
          dprime_before(i)=fmins('FitdvPercent', init_dprime, [], [],per_correct_before(i), task);
          dprime_after(i)=fmins('FitdvPercent',  init_dprime, [],[],per_correct_after(i), task);
else
          dprime_before(i)=fminsearch('FitdvPercent', init_dprime,[],per_correct_before(i), task);
          dprime_after(i)=fminsearch('FitdvPercent',  init_dprime, [],per_correct_after(i), task);
end
end

% find the contrast for which dprime=0.5 before training
init_dprime=0.5;
interp_contrast=interp1(dprime_before, contrast, init_dprime); %the contrast for which d'==.5

% find the dprime value after training for the contrast at which dprime=0.5
% before training).
new_dprime=interp1(contrast, dprime_after, interp_contrast);

subplot(1, 2, 1)
plot(contrast, per_correct_before, 'k', contrast, per_correct_after, 'k--');
xlabel('contrast')
ylabel('percent correct')
legend('before training', 'after training')
```

```
subplot(1, 2, 2)
plot(contrast, dprime_before, 'k', contrast, dprime_after, 'k--');
xlabel('contrast')
ylabel('dprime')


%*****************************************%
function L=FitdvsPercent(dprime, correct, task);

% finds the d prime separation for a given percent correct
% uses maximum likelihood function minimization
bestper=dvPercent(dprime,task);
L=(correct-bestper)^2;


%*****************************************%
function bestper=dvPercent(dprime,task);

% finds the percent correct (assumining an optimal criterion etc.) for a given
% dprime

% creates signal and noise distributions, assuming signal and noise
% have equal standard deviations of 1
% distributions are scaled by the standard deviation of the noise
sigS=1; %standard deviation of signal
sigN=1; %standard deviation of noise
x=linspace(-10, 10, 1000)./sigN;
dprime=dprime/sigN;
sigS=sigS/sigN;

%calculate the hit/false alarm rate

hit=1-NormalCumulative(x, dprime, sigS^2);
fa=1-NormalCumulative(x, 0, sigS^2);

if (task=='YESN') %yes-no
      correctvals=hit+(1-fa); %assuming the criterion is that yes and no equally likely
      beta=find(correctvals==max(correctvals));
      bestper=hit(beta(1));

elseif (task=='2ALT') %2-alt FC
      bestper=0;
      bestper=sum((hit(1:length(x)-1)-hit(2:length(x))).*(1-fa(2:length(x))));

elseif (task=='3ALT')
      bestper=0;
      bestper=sum((hit(1:length(x)-1)-hit(2:length(x))).*((1-fa(2:length(x))).^2));

elseif (task=='4ALT')
      bestper=0;
      bestper=sum((hit(1:length(x)-1)-hit(2:length(x))).*((1-fa(2:length(x))).^3));

elseif (task=='SMDF')% same different
      correct1=hit+(1-fa);%assuming the criterion is that same and different equally likely
      beta=find(correct1==max(correct1));
      beta=beta(1);
      bestper=2*(hit(beta))^2-2*hit(beta)+1;
end


%*****************************************%
function prob = NormalCumulative(x,u,var)

% function prob = NormalCumulative(x,u,var)
% Compute the probability that a draw from a N(u,var)
% distribution is less than x.
% Taken from the psychophysics toolbox
% http://www.psychtoolbox.org//
% 6/25/96  dhb  Fixed for new erf convention.

[m,n] = size(x);
z = (x - u*ones(m,n))/sqrt(var);
prob = 0.5 + erf(z/sqrt(2))/2;
```

Acknowledgments

References

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research, 36,* 3487-3500. [PubMed]

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science, 218,* 697-698. [PubMed]

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27,* 953-965. [PubMed]

Beard, B. L., Levi, D. M., & Reich, L. N. (1995). Perceptual learning in parafoveal vision. *Vision Research, 35,* 1679-1690. [PubMed]

Bennett, R. G., & Westheimer, G. (1991). The effect of training on visual alignment discrimination and grating resolution. *Perception and Psychophysics, 49,* 541-546. [PubMed]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433-436.

Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *Journal of Neurophysiology, 78,* 2889-2894. [PubMed]

Crist, R. E., Li, W., & Gilbert, C. D. (2001). Learning to see: Experience and attention in primary visual cortex. *Nature Neuroscience, 4,* 519-525. [PubMed]

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4,* 2051-2062. [PubMed]

Dorais, A., & Sagi, D. (1997). Contrast masking effects change with practice. *Vision Research, 37,* 1725-1733. [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, U S A, 95,* 13988-13993.

Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research, 39,* 3197-3221. [PubMed]

Fahle, M., & Edelman, S. (1993). Long-term learning in Vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research, 33,* 397-412. [PubMed]

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research, 35,* 3003-3013. [PubMed]

Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research, 23,* 145-150. [PubMed]

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature, 287,* 43-44. [PubMed]

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vision Research, 21,* 1149-1158. [PubMed]

Fine, I., & Jacobs, R. A. (2000). Perceptual learning for a pattern discrimination task. *Vision Research, 40,* 3209-3230. [PubMed]

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: Object specificity and size invariance. *Vision Research, 40,* 473-484. [PubMed]

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences, U S A, 96,* 3314-3319.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38,* 2401-2428. [PubMed]

Ghose, G. M., Yang, T., & Maunsell, J. H. (2002). Physiological correlates of perceptual learning in monkey v1 and v2. *J Neurophysiol, 87(4),* 1867-1888. [PubMed]

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature, 402,* 176-178. [PubMed]

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a Vernier discrimination task. *Vision Research, 37,* 2133-2141. [PubMed]

Johnson, C. A., & Leibowitz, H. W. (1979). Practice effects for visual resolution in the periphery. *Perception and Psychophysics, 25,* 439-442. [PubMed]

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science, 265,* 603-604. [PubMed]

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology, 80,* 324-330. [PubMed]

Liu, A., & Vaina, L. M. (1998). Simultaneous learning of motion discrimination in two directions. *Cognitive Brain Research, 6,* 347-349. [PubMed]

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5,* 552-563. [PubMed]

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge University Press.

Matthews, N., Liu, Z., Geesaman, B. J., & Qian, N. (1999). Perceptual learning on orientation and direction discrimination. *Vision Research, 39,* 3692-3701. [PubMed]

Matthews, N., & Welch, L. (1997). Velocity-dependent improvements in single-dot direction discrimination. *Perception and Psychophysics, 59,* 60-72. [PubMed]

Mayer, M. J. (1983). Practice improves adults' sensitivity to diagonals. *Vision Research, 23,* 547-550. [PubMed]

McKee, S. P., & Westheimer, G. (1978). Improvement in Vernier acuity with practice. *Perception and Psychophysics, 24,* 258-262. [PubMed]

Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision, 10,* 51-58. [PubMed]

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437-442. [PubMed]

Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology, 73,* 713-726. [PubMed]

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). Practising orientation identification improves orientation coding in V1 neurons. *Nature, 412,* 549-553. [PubMed]

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics, 52,* 582-588. [PubMed]

Sigman, M., & Gilbert, C. D. (2000). Learning to find a shape. *Nature Neuroscience, 3,* 264-269. [PubMed]

Smallman, H. S., MacLeod, D. I. A., & Doyle, P. (2001). Realignment of cones after cataract removal. *Nature, 412,* 604-605. [PubMed]

Swets, J.A. (1964) *Signal detection and recognition by human observers: Contemporary readings.* New York, Wiley.

Treue, S., & Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature, 382,* 539-541. [PubMed]

Vidyasagar, T. R., & Stuart, G. W. (1993). Perceptual learning in seeing form from motion. *Proceedings of the Royal Society of London, B, 243,* 241-244.

Vogels, R., & Orban, G. (1985). The effect of practice on the oblique effect in line orientation judgements. *Vision Research, 11,* 1679-1689.

Westheimer, G. (2001). Is peripheral visual acuity susceptible to perceptual learning in the adult? *Vision Research, 41,* 47-52. [PubMed]

Young, M. P., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science, 256,* 1327-1331. [PubMed]

Zohary, E., Celebrini, S., Britten, K. H., & Newsome, W. T. (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science, 263,* 1289-1292. [PubMed]