

Learning to detect and combine the features of an object

Jordan W. Suchow^a and Denis G. Pelli^{b,1}

^aDepartment of Psychology, Harvard University, Cambridge, MA 02138; and ^bDepartment of Psychology and Center for Neural Science, New York University, New York, NY 10003

Edited by Wilson S. Geisler, The University of Texas at Austin, Austin, TX, and approved November 19, 2012 (received for review October 23, 2012)

To recognize an object, it is widely supposed that we first detect and then combine its features. Familiar objects are recognized effortlessly, but unfamiliar objects—like new faces or foreign-language letters—are hard to distinguish and must be learned through practice. Here, we describe a method that separates detection and combination and reveals how each improves as the observer learns. We dissociate the steps by two independent manipulations: For each step, we do or do not provide a bionic crutch that performs it optimally. Thus, the two steps may be performed solely by the human, solely by the crutches, or cooperatively, when the human takes one step and a crutch takes the other. The crutches reveal a double dissociation between detecting and combining. Relative to the two-step ideal, the human observer's overall efficiency for unconstrained identification equals the product of the efficiencies with which the human performs the steps separately. The two-step strategy is inefficient: Constraining the ideal to take two steps roughly halves its identification efficiency. In contrast, we find that humans constrained to take two steps perform just as well as when unconstrained, which suggests that they normally take two steps. Measuring threshold contrast (the faintness of a barely identifiable letter) as it improves with practice, we find that detection is inefficient and learned slowly. Combining is learned at a rate that is 4× higher and, after 1,000 trials, 7× more efficient. This difference explains much of the diversity of rates reported in perceptual learning studies, including effects of complexity and familiarity.


object recognition | sensitivity | letter identification

The world is full of objects, and we spend our lives identifying them. Reading an hour a day for a year means identifying millions of letters and words. Each letter is a good basic-level object: simple, common, useful, and with its own name and shape (1–4). Identifying a letter requires two steps of visual processing: the observer first detects the letter's features and then combines them to recognize the letter (5).

However, what is a feature? Interpretation of learning studies that use traditional letters and other everyday objects is hindered by the infinite number of possible features, which include physical properties, like size and shape, as well as abstract properties, like function and beauty (6–8). To avoid this morass, we narrowly define features as discrete components of an image that are detected independently of each other (5).

When letters share features (perhaps, the vertical bar in a D and an L), detecting one feature is not always enough to tell which letter it is, so multiple features must be detected and combined for reliable identification. Both steps—detection and combination—are liable to errors that impede identification. For example, if a letter is faint or seen in dim light, a reader may incorrectly identify it because she fails to detect a feature that is present or because she spuriously “detects” a feature that is absent. Identifying an unfamiliar letter can be difficult even when all of its features have been correctly detected. For example, a novice reader may mistakenly identify a plainly visible letter, confusing the shape of one for that of another.

Whether struggling to detect or to combine, with more practice, observers fail less. They learn. Feature detection and combination can both be learned through practice (9–11).

To study features, it is helpful to use Gabors. A Gabor is a grating patch that is made by vignetted a sinusoidal grating with a Gaussian window, which restricts its spatial extent to a few bars . Gabors are fairly well matched to the receptive fields of simple cells in the primary visual cortex measured physiologically, and to the tuning of spatial frequency channels measured psychophysically. Gabors can differ in position, orientation, and spatial frequency. If Gabors are sufficiently different along these dimensions, they are detected independently and can be distinguished by a single feature detection (12, 13). Practice improves detection of a Gabor (14). This learning is specific to the trained stimulus and location (15, 16).

Tasks requiring feature combination also improve with practice. Merely detecting the presence of an object does not require combining its features, but identifying it usually does; this is because detecting any feature reveals the object's presence, but, depending on the other possible objects, usually several features are needed to specify which object is present. Fine and Jacobs (17) measured improvement with practice in identifying compound gratings, which are multifeature objects composed of several superimposed Gabors, and found that learning transferred across orientations, unlike learning in detection tasks. Likewise, Kovács et al. (18) measured improvement of search for orientation-defined contours and found that learning transferred between eyes and to other orientation-defined contours, again unlike learning in detection tasks.

There are hints that the two steps, detection and combination, may be learned at different rates. Learning of familiar letters is slow and has been attributed to improved feature detection (19). Unlike the slow learning of familiar letters, the learning of new letters is initially fast, but slows as the letters become familiar (5, 20–22). This learning might involve improvement at either step. Identification involves both detecting and combining of features, so, when identification performance improves, one would like to know how much of this learning is due to improved detection rather than improved combination of features.

Here, through the use of six variously enhanced observers performing the same letter-identification task, we dissociate detecting and combining, revealing each step's contribution to learning. Of the six kinds of observer, two are “unconstrained” and four are “composite.” Unconstrained is the traditional situation of presenting a faint target and asking the observer to identify it, with no constraints. We test both the human (H) and the ideal (I) observer in this way. The ideal is an algorithm that chooses the most probable hypothesis, maximizing expected accuracy. Composite observers are new. Of the four composite observers, two are bionic. They are human in only one of the steps. The other step is delegated to a bionic crutch, either the ideal detector or the ideal combiner. In these two cases, the two perform as a team:

Author contributions: J.W.S. and D.G.P. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: denis.pelli@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1218438110/-DCSupplemental.

Either the human combines what the ideal detects (composite IH), or the ideal combines what the human detects (composite HI). Having broken up the task into two parts, we can also assign both parts, in distinct sessions, to the same observer, so that the human (composite HH) or the ideal (composite II) takes both steps. The bionic crutches test for double dissociation: Is the human identification process actually separable into two distinct steps of detection and identification?

Are the bionic crutches overkill? Is it not enough, for our purpose, just to note the different learning rates for tasks that do and do not require combining? No. That comparison is suggestive, but has not led to any published conclusions about distinct learning rates in separate stages. Each task yields a rate, but no one has managed to link the task and the model strongly enough to draw conclusions that distinguish the two kinds of learning in one model. Adding to the confusion, much of the perceptual learning results are for fine discrimination of one feature, like orientation, which may require combining the activity of several feature detectors, but has usually been taken to reflect learning at an early stage. Our bionic crutches provide a double-dissociation paradigm that rises above these vagaries, showing how the observed pattern of results is diagnostic evidence for independent processes. Our paradigm requires manipulations (the bionic crutches) that selectively affect the two presumed processes. The strong conclusion is well worth the bother.

To separate the steps, we need to know the letters' features; they are uncertain for traditional letters, so we use Gabor letters instead (Fig. 1). Based on the probability summation literature, we suppose that our Gabors are features, detected independently (12, 13). The juxtaposition of n Gabors creates an n -feature "letter" (23, 24). Incidentally, though Gabors are very well-suited to be the elements of our stimuli, they are not essential; simple bars might do as well. By using Gabor (or bar) letters, we can precisely specify the features that constitute each letter, while maintaining the essence of an alphabet: a set of many distinguishable objects sharing a common visual style. Our Gabor letters are similar to Braille letters in that they each consist of a binary array of features. Braille behaves well when presented visually (5, 25). Even so, the conclusions of this paper do not depend on our claim that Gabor "letters" are letters; it is enough that they are objects.

We created the IndyEighteen alphabet. In general, an Indy N alphabet is the set of all possible combinations of N features. Suppose we are asked to identify a randomly selected letter from



Fig. 1. Eight Gabor letters. The letters of the IndyEighteen alphabet are composed of Gabors. Each of the 18 possible Gabors is oriented $\pm 45^\circ$ from vertical and is at one of nine locations in a 3×3 grid. When a right-tilted and a left-tilted Gabor coincide, they form a plaid, but vision still responds to them independently. We suppose that the Gabors are detected independently, so that each Gabor is a feature. With two orientations and nine locations, there are 18 possible Gabors, i.e., features. The eight letters displayed here are a randomly selected subset of the 2^{18} letters in the whole alphabet. Note that within this subset, some features are common to many letters (e.g., six of the eight letters contain a right-tilted Gabor at the top right corner), whereas some features are common to just a few (e.g., two of the eight letters contain a right-tilted Gabor at the bottom left position).

this alphabet of 2^N letters. Because the presence of each feature is statistically independent of the rest, all N features must be detected to identify the letter reliably. In most traditional alphabets, however, a letter can be identified without detecting all of its features.* To better match this property of traditional alphabets, we created several eight-letter subsets drawn randomly from IndyEighteen. One such subset appears in Fig. 1. Reducing the number of possible letters makes identification easier. In general, in a subset of Indy, the features are no longer independent or equally frequent, so fewer feature detections are needed for identification, and some features are less informative than others. At the extremes, a feature may be unique to a letter and thus diagnostic of its identity, or common to all of the letters and thus irrelevant to the task of distinguishing among them (26).[†]

For each unconstrained or composite observer, we create a new alphabet consisting of eight IndyEighteen letters. On each trial, we ask the observer to identify a letter drawn from that eight-letter alphabet. We measure threshold contrast, the lowest contrast (faintness) sufficient to identify the letters correctly 75% of the time. We then convert threshold contrast to efficiency. Efficiency is a useful way to characterize performance of a computational task (27, 28); this pits the actual observer against the ideal observer, an algorithm that performs the whole task optimally, not constrained to taking two steps. Efficiency is defined as the fraction of the signal energy used by an observer that is required by the ideal to perform just as well. Contrast energy is proportional to the contrast squared, so the efficiency of the actual observer is

$$\eta = \frac{c_1^2}{c^2}, \quad [1]$$

where c and c_1 are threshold contrasts of the actual and ideal observers.

Results

Dissociating Detecting from Combining. Fig. 2 shows learning for two participants, plotting threshold contrast as a function of the number of completed identification trials. (Results for all six participants appear in Fig. S1.) The right-hand vertical scale shows the efficiency corresponding to each threshold contrast. There are two graphs (Fig. 2, *Left* and *Right*), one per participant. Within each graph appear all results for that participant, unconstrained and composite. The top line (Fig. 2, solid black line) is the unconstrained ideal (I), the baseline for calculating efficiency. The bottom solid line is the unconstrained human. The other four lines, sandwiched in between, are for composites. Solid lines are fits to data, and the dashed line is a prediction derived from the other lines (Eq. 3). The vertical positions of the lines show that threshold contrast (and efficiency) are best for the unconstrained ideal, slightly worse for the two bionic crutches working together, and get worse, from line to line, as we ask the human to do part or all of the work (Fig. 2, bottom solid line). At trial 1,000, the composite-observer efficiency with the human doing just the combining (IH, 15%) is 7 \times that with the human doing just the detecting (HI, 2.1%). The lines in which the human does

*Pelli et al. (5) found that human observers need 7 ± 2 feature detections for threshold letter identification for all traditional alphabets tested, over a 10-fold range of complexity. Assuming that feature count is proportional to complexity, as proposed in ref. 5, then, even if the least-complex alphabet tested had only seven features per letter, the most complex had 70 features per letter. Thus the seven features detected at the threshold for identification of a complex letter are only a small fraction of the letter's features.

[†]Using a Monte Carlo simulation, we determined that 4–14 feature detections are required to achieve criterion performance of 75% correct for identifying a letter from a set of eight randomly selected IndyEighteen letters, depending on the false alarm rate. A false alarm occurs when an absent feature is "detected." Sometimes, by chance, enough features are falsely detected such that the letter appears more similar to one of the foils than to the target. Additional feature detections, hits, are needed to compensate. We considered false alarm rates between 0% and 51%. At false alarm rates greater than 51%, it is impossible for the observer to achieve criterion performance, even with a hit rate of 100%.

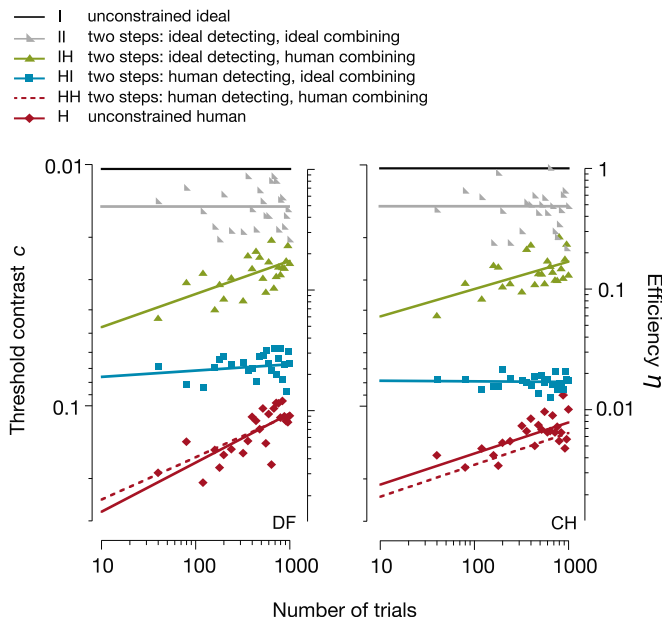


Fig. 2. Learning. The participant (DF, *Left*; CH, *Right*) detects and combines features to identify a letter from an alphabet of eight different IndyEighteen letters. Each unconstrained or composite observer trained at threshold contrast, receiving just enough contrast to achieve criterion performance (75% correct). Each line shows an unconstrained or composite observer's threshold contrast c (using the contrast scale, *Left*) and efficiency η (using the efficiency scale, *Right*). The bottom solid line is the unconstrained human (H), and the top line is the unconstrained ideal (I). The dashed line is the composite human (HH). Separability of the two steps predicts that the unconstrained and composite observers will perform equally, which is approximately true for the human H and HH (bottom two lines, solid and dashed) and is not true of the ideal I II (top two lines). The horizontal scale counts identification trials. For composite HI, the human performs 18 detection trials for each identification trial. In both vertical scales, learning goes up: efficiency increases upward and threshold contrast increases downward. Because efficiency is inversely proportional to threshold contrast squared (Eq. 1), the log-log slope of efficiency is $-2\times$ that of threshold contrast.

the combining (H, HH, and IH) are steep, showing fast learning, and the rest (I, II, HI) are shallow. The log-log slope of pure combination learning (IH, -0.11 ± 0.01 ; Fig. 2, green triangles) is $4\times$ that for pure detection learning (HI, -0.03 ± 0.01 , blue squares). The reported slopes and efficiencies are averages across all participants. (As a control, four of the participants used a modified version of the bionic crutch with independent detection trials, as described in *Supporting Materials and Methods*, Composite HI'.)

Does the unconstrained human really take two steps, first detecting and then combining? That is an inefficient way to identify. Constraining the ideal to take two steps roughly halves its efficiency, $\eta_{II} \sim 0.5\eta_I$, resulting in the gap between the upper two lines in Fig. 2. However, constraining humans to take two steps leaves their efficiency unchanged, $\eta_{HH} \sim \eta_H$. This is the coincidence of the dashed and solid lines at the bottom of Fig. 2, which do not differ significantly from each other in slope or intercept across participants (paired t test, $P = 0.15$ and $P = 0.92$, respectively). Forcing people to take two steps does not impair their performance, which suggests that taking two steps may be an intrinsic limitation of human object recognition.

Do the bionic crutches really isolate the contributions to human performance of two distinct processes? More precisely, does each crutch boost efficiency by a multiplicative factor (Eq. 2)? In short, are the steps separable? That conjecture is tested and verified by the agreement of the two-step and unconstrained human performance (H and HH, dashed and solid lines at the bottom of Fig. 2). Thus, we have dissociated the contribution of each step (29); in

the language of dissociation studies, this is a within-task process decomposition with a multiplicative composite measure: efficiency. The task is letter identification; the provision of each bionic crutch—detector or combiner—is an independent manipulation. Finding this double dissociation of detecting and combining shows that object recognition “is accomplished by a complex process that contains two functionally distinct and separately modifiable parts” (ref. 29, p. 180).

The bionic crutch paradigm can be applied to any observer, biological or not. However, the results are particularly easy to interpret when a double dissociation is revealed, as we found for the humans, but not for the ideal: the human observer's internal computation really is separable into detection and combination steps, with the overall efficiency equal to the product of the efficiencies of the steps (Eq. 3).

Efficiency for identifying a letter or a word is inversely proportional to complexity or word length (3, 5). Wondering why, it has seemed obvious that the low overall human efficiency for identifying a word (e.g., 1% for a short word) is mostly due to the two-step strategy that detects the parts independently before combining for identification. Supposing that vision is mediated by feature detection seems to imply that each feature must reach threshold by itself, and this applies equally to the letters in a word and the features in a letter. Thus, it once seemed to us that humans are inefficient mostly because of the inefficiency of taking two steps. With this background, perhaps the reader will share our surprise in discovering, through Monte Carlo trials, that the ideal two-step strategy has a respectable efficiency of at least 65% for any number of parts. We are astounded that the cost of reducing each feature's sensory information to a bit has such a modest effect on overall efficiency. Thus, the mere fact of taking two steps does not doom the combining efficiency to fall inversely with complexity, as the human's does. Perhaps the human drop in combining efficiency is due to a limit in the number of features that the observer combines, say 7 ± 2 , as indicated by the ratio of thresholds for identification and detection (5).[‡]

Comparing Slopes to Explain Effects of Familiarity and Complexity on Learning.

The shallow slope of detection learning, over 1,000 trials, matches that of other detection-learning studies, which show slow learning over many sessions. Twelve slopes of learning curves, from this study and seven other published papers, are displayed in Table 1 and Fig. 3 (5, 13, 16, 21, 30–32). We limit our survey to studies that report threshold contrasts for tasks that demand discrimination of objects and patterns presented in central vision, at fixation.



Our result—the shallow slope of learning to detect and the steep slope of learning to combine—can explain the effects of stimulus complexity and familiarity on the rate of learning, where complex objects (requiring discrimination along many perceptual dimensions) are learned faster than (simple) Gabors, and where unfamiliar objects are learned faster than familiar objects.











It seems that complex objects are learned more quickly than Gabors because complex objects require combining. Comparing across studies in the literature (Table 1), we find that learning to identify stimuli that require combining, such as unfamiliar faces (slope $b = -0.40$), bandpass-filtered noise textures (-0.26), 4×4 random-checkerboard patterns (-0.16), and compound gratings (-0.21), is much quicker than learning to detect a Gabor (-0.03 , -0.06), which does not require combining.

However, combination learning soon saturates, as the letters become familiar. Extrapolating the fitted line for human combination (Fig. 2, IH) predicts that efficiency would reach 100% (ideal combining) after 1 million trials. Typical readers read a million letters every 2 wk, for years. With so much experience,

[‡]Vul E, Goodman ND, Griffiths TL, Tenenbaum JB (2009) One and done? Optimal decisions from very few samples. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds Taatgen NA, van Rijn H (Cogn Sci Soc, Austin, TX), July 29, 2009, pp 66–72.

Table 1. Slope of detection and identification learning in various experiments reported here and in seven other published papers

Stimuli	Slope b	Familiar	Source
 Gabor (detection)	-0.03	Yes	 Human detects and ideal combines
 Familiar letter	-0.04	Yes	Pelli et al. (5)
 Gabor (detection)	-0.06	Yes	Furmanski et al. (14)

 Gabor letter	-0.11	No	 Ideal detects and human combines
 Unfamiliar letters	-0.11	No	Suchow and Pelli (22)
 4 × 4 random checkerboard	-0.11	No	Pelli et al. (5)
 Gabor (fine discrimination)	-0.12	No	Lu and Doshier (30)
 Gabor letter	-0.16	No	 Human, unconstrained
 Filtered noise texture	-0.32	No	Gold et al. (31)
 Unfamiliar face	-0.36	No	Gold et al. (31)
 Shape in filtered noise	-0.48	No	Michel and Jacobs (32)
 Compound grating	-0.78	No	Fine and Jacobs (17)

Plotting threshold contrast as a function of the number of completed trials, we fit parameter b , the log-log slope. The dashed line separates the familiar from the unfamiliar.

surely they have learned to combine as well as they can. Any additional learning of these familiar letters likely occurs in the detection step. This is presumably why the slope of learning familiar letters (-0.02) matches our measured slope of learning in the detection step (-0.03). When both the task and stimuli are familiar (e.g., identifying a familiar letter), the slope of learning falls on one side of the dashed line, showing slow learning (Table 1 and Fig. 3). Slopes of learning unfamiliar tasks or stimuli fall on the other side. Presumably the steep slope for unfamiliar stimuli is the fast learning of combination, which saturates once the stimuli are familiar, leaving only the slow learning of detection.

Number of Features and Extent of Each Feature. We find that after 1,000 trials with an eight-letter subset of Indy18, the 15% combination efficiency (IH) is 7× the 2.1% detection efficiency (HI). In two-stage identification, each feature is detected independently, so we expect the detection efficiency to be independent of the number of features. The Gabor that we used as a feature was fairly extended. Detection efficiency could be raised by using a less-extended Gabor, with fewer bars. Because HI and II efficiencies are nearly independent of the number of features, and H and HH efficiencies are inversely proportional to the number of features, Eq. 3 implies that combination efficiency must be inversely proportional to the number of features, which could be explored by testing Indy4 and Indy100, say. Thus, reducing the number of features would increase combining efficiency without affecting the detection efficiency. Reducing the Gabor extent would increase the detection efficiency without affecting the combining efficiency.

Beyond Gabor Letters. It may be possible to extend our approach beyond Gabor letters to other stimuli, such as words, faces, and

scenes, whose features are unknown. If one assumes the separability found here, then it may be easy to factor out the efficiency of detecting (Eq. 5). Alternatively, mild image transformations, like scaling and translation, change the features but preserve abstract properties of the feature combination, like shape, that may determine the object's identity. We noted at the outset that the existing literature on perceptual learning in early and late visual processes suggests that combination learning transfers across mild transformations and detection learning does not. In human observers, the steps are separable: Overall composite efficiency is the product of the composite efficiencies of the two steps (Eq. 6). Thus, for identification of an object from an arbitrary set, measuring the partial transfer of learning across a mild transformation like scaling or translation would distinguish the contributions of both steps: feature detection and combination.

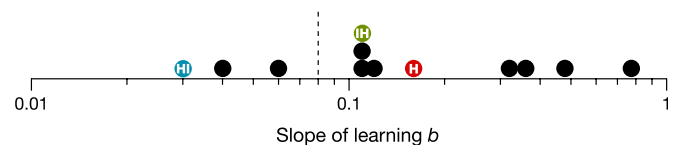





Fig. 3. Slope of learning. Histogram of the slopes of learning in Table 1, with one symbol for each row of the table. The labeled symbols , , and  represent our identification tasks. The horizontal position of each symbol is the log-log slope b . The dashed vertical line corresponds to the dashed horizontal line in Table 1. This histogram shows the dichotomy of fast learning of unfamiliar objects and slow learning of familiar objects. We speculate that the fast learning of unfamiliar objects is learning to combine (i.e., recognize the shapes), which quickly saturates, such that, once those objects have become familiar, we are reduced to learning slowly as we gradually learn to detect the features better.

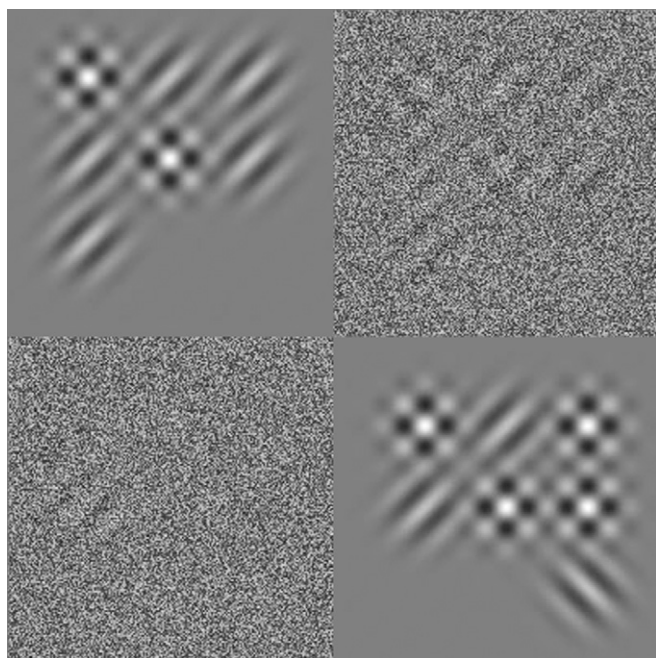


Fig. 4. Stimuli. (*Upper Left*) A Gabor letter. When unconstrained, the human participant is presented with a Gabor letter faintly in noise (*Upper Right*). As the detector, the participant is presented with a single feature faintly in noise (*Lower Left*) and, as the combiner, with an imperfect set of detected features (*Lower Right*). In this last case, the high-contrast Gabors are easily seen, but are a less-than-faithful copy of the original letter's features, which makes it hard to guess what the original letter was.

Materials and Methods

On each trial, we ask the unconstrained or composite observer to identify an IndyEighteen letter in added white noise. The letter and noise are both static, presented together for 200 ms. Testing of each unconstrained or composite observer begins with a new eight-letter alphabet and is performed in a single block of 25 runs, with 40 trials per run. Short (2-min) breaks are taken between runs, as needed, and longer (30-min) breaks are taken between blocks. The entire session was completed within 8 h in 1 d, without sleep or naps. The order of the blocks (one per task) is randomized for each observer to minimize any order effect in the group average.

Unconstrained H: Human Identifies. The human participant identifies, unconstrained. On each trial, we present a letter at threshold contrast (Fig. 4) to the human participant, who identifies the letter by selecting it from the response screen (Fig. 1). This trial challenges the human to identify, presumably by detecting and combining.

Unconstrained I: Ideal Identifies. The ideal observer identifies, unconstrained. The human participant plays no role. On each trial, we present a letter at threshold contrast. The ideal identifies the letter by choosing the most likely possibility; it compares the noisy stimulus to each letter on the response screen at the contrast of the signal, and selects the most similar (minimum rmsd; see appendix A of ref. 5). The ideal achieves the best possible expected performance, and this is the baseline for calculating efficiency.

Composite HI: Two Steps (Human Detects, Ideal Combines). The human participant detects and the bionic crutch (ideal combiner) combines. On each identification trial, instead of being shown the whole letter in a single presentation, the human performs 18 detection trials, one for each possible feature. (The 18 detection trials count as one identification trial in the horizontal axis of Fig. 2.) On each detection trial, the human participant reports whether the feature is present by responding "present" or "absent." The 18 present-vs.-absent decisions are recorded as an 18-bit string (1 if present; 0 if otherwise) that is passed to the bionic crutch (ideal combiner). The ideal combiner makes its selection by comparing the string received to the string for each letter on the response screen, selecting the most similar

(minimum Hamming distance) (33); this challenges the human participant to detect, without challenging combination.

Composite IH: Two Steps (Ideal Detects, Human Combines). The bionic crutch (ideal detector) detects and the human participant combines. On each identification trial, the crutch performs 18 detection trials. On each detection trial, the crutch selects the most probable hypothesis (present or absent), given the noisy stimulus and the frequency of that feature in the alphabet. Features judged by the crutch to be present are displayed at high contrast to the human participant, who identifies the letter by selecting it from the response screen; this challenges the human to combine, without challenging detection.

Composite II: Two Steps (Ideal Detects and Combines). The two bionic crutches together perform the whole task, in cascade. The human participant plays no role.

Composite HH: Two Steps (Human Detects and Combines). The human participant takes the two steps in separate sessions, one for each step. This trial challenges the human to detect in one session, and to combine in another session. The level of performance achieved by this two-step composite observer, HH, is computed from the measured performance of the other three two-step composites: HI, IH, and II. For this calculation, we suppose that the efficiency η of each two-step composite observer is the product of two factors, a and b , one for each step, and that each factor depends on whether that step is performed by the human (H) or an ideal bionic crutch (I), but it is independent of how the other step is performed:

$$\begin{aligned}\eta_{HH} &= a_H b_H \\ \eta_{HI} &= a_H b_I \\ \eta_{IH} &= a_I b_H \\ \eta_{II} &= a_I b_I.\end{aligned}\quad [2]$$

Because multiplication is transitive, we easily solve for the two-step human efficiency in terms of the others:

$$\eta_{HH} = \eta_{HI} \eta_{IH} / \eta_{II}.\quad [3]$$

This equation can be recast as a statement about thresholds, using Eq. 1 to substitute thresholds for efficiencies,

$$c_{HH} = c_{HI} c_{IH} / c_{II}.\quad [4]$$

Both equations correspond to the same dashed line in Fig. 2, using the threshold scale on the left (Eq. 4) or the efficiency scale on the right (Eq. 3). In future work, it will be interesting to study the human combination efficiency η_{HI} , for which we can solve Eq. 3,

$$\eta_{HI} = \eta_{HH} \eta_{II} / \eta_{IH}.\quad [5]$$

All of the terms on the right of Eq. 5 are easily accessible. η_H is easy to measure, and our work here suggests that, in future studies, one might assume that $\eta_{HH} = \eta_H$. The human efficiency of detecting η_{HI} seems to be conserved across many conditions, so that it could be estimated once. And the two-step efficiency η_{II} is easily computed by implementing the one- and two-step ideals. In this way, Eq. 5 could make it easy to routinely estimate the observer's combining efficiency η_{IH} .

Eq. 3 may seem odd if you did not expect the η_{II} term there for two-step efficiency; we can make it more intuitive by defining composite efficiency η_j relative to the composite ideal, II. Recall that standard efficiency is $\eta = E_i/E$. We now define composite efficiency $\eta_j = E_{ij}/E$. In this new notation, Eq. 3 becomes

$$\eta_{HH} = \eta_{HI} \eta_{IH}.\quad [6]$$

In words, for any observer whose efficiency is separable (Eq. 2), the overall composite efficiency is the product of the composite efficiencies of the steps. Eqs. 2–6 are all equivalent. Though the equation for η_j (Eq. 6) is simpler and more intuitive than the equation for η (Eq. 3), we chose to plot the traditional familiar efficiency η rather than our new-fangled composite efficiency η_j because they differ solely by the factor η_{II} , which is nearly 1.

ACKNOWLEDGMENTS. We thank Chris Berendes, Y-Lan Boureau, Charles Bigelow, Rama Chakravarthi, Hannes Famira, Judy Fan, Jeremy Freeman, Ariella Katz, Yann LeCun (isolating detection), Christine Looser, Najib Majaj,

Charvy Narain, Robert Rehder, Wendy Schnebelen, Elizabeth Segal, Eva Suchow, Steven Suchow, Katharine Tillman, Bosco Tjan (adding the unconstrained ideal), and Ed Vessel for helpful comments and discussion. We thank several

anonymous reviewers for many helpful suggestions. This is draft 146. This research was supported by National Institutes of Health Grant R01-EY04432 (to D.G.P.).

1. Rosch E, Mervis CB, Gray W, Johnson DM, Boyes-Braem P (1976) Basic objects in natural categories. *Cognit Psychol* 8:382–439.
2. Wong ACN, Gauthier I (2007) An analysis of letter expertise in a levels-of-categorization framework. *Vis Cogn* 15:854–879.
3. Pelli DG, Farell B, Moore DC (2003) The remarkable inefficiency of word recognition. *Nature* 423(6941):752–756.
4. Pelli DG, et al. (2009) Grouping in object recognition: The role of a Gestalt law in letter identification. *Cogn Neuropsychol* 26(1):36–49.
5. Pelli DG, Burns CW, Farell B, Moore-Page DC (2006) Feature detection and letter identification. *Vision Res* 46(28):4646–4674.
6. Treisman A (1988) Features and objects: The fourteenth Bartlett memorial lecture. *Q J Exp Psychol A* 40(2):201–237.
7. Pinker S (1984) Visual cognition: An introduction. *Cognition* 18(1-3):1–63.
8. Murphy GL (2002) *The Big Book of Concepts* (MIT Press, Cambridge, MA).
9. Gibson E (1969) *Principles of Perceptual Learning and Development* (Appleton-Century-Crofts, New York).
10. Fine I, Jacobs RA (2002) Comparing perceptual learning tasks: A review. *J Vis* 2(2):190–203.
11. Ahissar M, Hochstein S (1997) Task difficulty and the specificity of perceptual learning. *Nature* 387(6631):401–406.
12. Watson AB (1979) Probability summation over time. *Vision Res* 19(5):515–522.
13. Robson JG, Graham N (1981) Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Res* 21(3):409–418.
14. Furmanski CS, Schluppeck D, Engel SA (2004) Learning strengthens the response of primary visual cortex to simple patterns. *Curr Biol* 14(7):573–578.
15. Mayer MJ (1983) Practice improves adults' sensitivity to diagonals. *Vision Res* 23(5):547–550.
16. Fahle M (2005) Perceptual learning: Specificity versus generalization. *Curr Opin Neurobiol* 15(2):154–160.
17. Fine I, Jacobs RA (2000) Perceptual learning for a pattern discrimination task. *Vision Res* 40(23):3209–3230.
18. Kovács I, Kozma P, Fehér A, Benedek G (1999) Late maturation of visual spatial integration in humans. *Proc Natl Acad Sci USA* 96(21):12204–12209.
19. Doshier BA, Lu ZL (1999) Mechanisms of perceptual learning. *Vision Res* 39(19):3197–3221.
20. Polk TA, Farah MJ (1995) Late experience alters vision. *Nature* 376(6542):648–649.
21. Chung ST, Levi DM, Tjan BS (2005) Learning letter identification in peripheral vision. *Vision Res* 45(11):1399–1412.
22. Suchow JW, Pelli DG (2005) Learning to identify letters: Generalization in high-level perceptual learning. *J Vis* 5(8):712, (abstr).
23. Levi DM, Hariharan S, Klein SA (2002) Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *J Vis* 2(2):167–177.
24. Levi DM, Sharma V, Klein SA (1997) Feature integration in pattern perception. *Proc Natl Acad Sci USA* 94(21):11742–11746.
25. Loomis JM (1981) On the tangibility of letters and braille. *Percept Psychophys* 29(1):37–46.
26. Seitz AR, Watanabe T (2009) The phenomenon of task-irrelevant perceptual learning. *Vision Res* 49(21):2604–2610.
27. Geisler WS (1989) Sequential ideal-observer analysis of visual discriminations. *Psychol Rev* 96(2):267–314.
28. Pelli DG, Farell B (1999) Why use noise? *J Opt Soc Am A Opt Image Sci Vis* 16(3):647–653.
29. Sternberg S (2003) Process decomposition from double dissociation of subprocesses. *Cortex* 39(1):180–182.
30. Lu ZL, Doshier BA (2004) Perceptual learning retunes the perceptual template in foveal orientation identification. *J Vis* 4(1):44–56.
31. Gold J, Bennett PJ, Sekuler AB (1999) Signal but not noise changes with perceptual learning. *Nature* 402(6758):176–178.
32. Michel MM, Jacobs RA (2008) Learning optimal integration of arbitrary features in a perceptual discrimination task. *J Vis* 8(2):3.1–16.
33. Hamming RW (1950) Error detecting and error correcting codes. *Bell Syst Tech J* 29(2):147–160.
34. Watson AB, Robson JG (1981) Discrimination at threshold: Labelled detectors in human vision. *Vision Res* 21(7):1115–1122.
35. Kim J, Wilson HR (1993) Dependence of plaid motion coherence on component grating directions. *Vision Res* 33(17):2479–2489.
36. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.
37. Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* 10(4):437–442.
38. Pelli DG, Zhang L (1991) Accurate control of contrast on microcomputer displays. *Vision Res* 31(7-8):1337–1350.
39. Watson AB, Pelli DG (1983) QUEST: A Bayesian adaptive psychometric method. *Percept Psychophys* 33(2):113–120.

Supporting Information

Suchow and Pelli 10.1073/pnas.1218438110

SI Materials and Methods

Composite II': Ruling Out Stimulus Artifacts. As a check on our work, we implemented composite II in two ways. First we created the composite observer by cascading the two bionic crutches that had already been programmed to work with the human, but now without the human. The ideal detector passes an 18-bit feature map, indicating which features were detected, to the ideal combiner. However, one reviewer noted that we create special stimuli, high-contrast letters, to test the human in composite IH, and perhaps some artifact in those stimuli is affecting our results. So we also implemented an ideal to identify the stimuli presented to the human in composite IH. This alternate implementation of composite II is formally equivalent, and gave identical results, assuring us that no stimulus artifact had intruded.

Composite HI': Ruling Out Unwanted Human Combination Learning in Composite HI. The whole point of the bionic observer with ideal combination is to isolate the human detection step, so it is essential to rule out combination learning by the human participant; we did that in three ways: by shuffling the order of detection trials to discourage pattern learning, by testing for it after training, and by running a modified control experiment that makes such learning impossible (i.e., a modified implementation of the bionic observer).

Distributing a letter's 18 possible features over 18 detection trials does not eliminate the letter's pattern; it merely converts a brief spatial pattern into a prolonged spatiotemporal pattern, extended over 18 detection trials. In principle, the observer could improve his detection performance by learning each letter's pattern, combining information across sequential presentations.

We discourage pattern learning by shuffling the order of the 18 detection trials that constitute each identification trial; in fact, the order is irrelevant, so this does not rule out pattern learning, but would likely hinder it. Furthermore, because shuffling makes it hard to guess the location of the next feature presentation, it has the incidental benefit of extending the relevant area to be attended to include the whole letter, not just one feature. Thus, the human participant is expected to attend to the whole letter area both when he detects as part of the bionic observer and when he identifies unaided.

At the end of training, we use a recognition test to discover any unwanted combination learning. On each trial, the human participant is shown two Gabor letters, one after the other, feature by feature, and is asked to indicate which of the two was used in training. As in training, on each trial, each letter's features are presented in random order. One of the letters is old, the other new. The old letter is a random sample from the training alphabet, an eight-letter subset of IndyEighteen. The new letter (the foil) is a random sample from a specially constructed alphabet, also an eight-letter subset of IndyEighteen, which has the same features as the old alphabet, rearranged into different combinations. This new alphabet is formed by shuffling the features of the eight old letters, across letters, to create the eight letters of the new alphabet. The two alphabets, old and new, have the same number of features of each type (first-order statistics), but differ in how these features are combined (higher-order statistics). Only if the observer has learned these combinations will he be able to distinguish old from new. No such learning was found: observers correctly indicated which letter was old on 46% of trials, not significantly different from chance (50%), or from performance measured before training, 52%.

To be absolutely sure of this key point, we also created a modified implementation of the bionic observer that eliminates the possibility of combination learning by the human detector.

Before, the 18 detection trials were all based on a single target letter. In the new implementation, there is still a target to be identified (by the ideal combiner), but each detection trial is based on an independently selected letter from the eight-letter alphabet, not necessarily the target. Each detection trial is conducted as a yes/no test, but scored as right or wrong. If the human detector is right, then the bionic combiner receives this feature correctly (i.e., present or absent), as in the target. If the human detector is wrong, then the combiner receives this feature wrongly (i.e., present if absent in the target; absent if present in the target). For the human, this abolishes the letter patterns (which are co-occurrences of features within a letter) while preserving the frequency of each feature in the alphabet as a whole; for the bionic combiner, this presents a feature vector reflecting human accuracy in detecting each feature.

Despite the several differences in our two implementations of detection, the outcome—separability—was the same for both the original and the variant, which suggests that the detection step is performed similarly, just as efficiently, in both cases. In the future, it will usually be enough to test for recognition after training to confirm that the pattern learning is negligible. If such a test reveals that combination learning has occurred, this modified implementation of the bionic observer with human detection and ideal combination can be used to eradicate it.

Participants. Six human participants (JS, AK, DF, CH, SS, and MC) performed unconstrained H. Four of them (DF, CH, SS, and MC) also performed in composites HI' and IH; the other two performed in composite HI. JS is an author. The others were naive to the purpose of the experiment. All participants gave informed consent in writing. Testing of human observers was approved by the NYU University Committee on Activities Involving Human Subjects (UCAIHS).

Stimuli. Signals are IndyEighteen letters. Each letter is a gray square with a combination of several Gabors oriented 45° from vertical, placed at various locations on a 3 × 3 grid. The superimposed Gabors are orthogonal, and, despite forming a plaid, they are detected and perceived independently (1, 2). The center-to-center spacing of adjacent Gabors is 1.4× the wavelength, 1.4/f. A vertical Gabor pattern is

$$L(x,y) = \left[1 + c \sin(2\pi fx) \exp\left(-\frac{x^2 + y^2}{\lambda^2}\right) \right] L_0,$$

where background luminance $L_0 = 21 \text{ cd/m}^2$, spatial frequency $f = 2 \text{ cycle/degree}$, spatial extent $\lambda = 0.61 \text{ degree}$, and contrast c is chosen using the estimate provided by QUEST.

Noise is added independently to each pixel of the stimulus, such that the luminance of any particular pixel is the sum of the luminance assigned to that pixel by the signal and a random increment or decrement in luminance sampled from a zero-mean Gaussian distribution, truncated at ± 2 SDs. The rms contrast of the noise is 0.20. There are 25.4 pixels/degree, horizontally and vertically. The power spectral density N is $10^{-4.21} \text{ deg}^2$.

Presentation. Stimuli are rendered by an Apple Macintosh computer running MATLAB in conjunction with the Psychophysics Toolbox extension (3, 4). Stimuli are displayed on a cathode-ray tube monitor, driving only the green gun to achieve 12-bit accuracy, at a background luminance of 21 candela/m² (5). The display resolution is set to 1,024 × 768 at 60 Hz, 29 pixels/cm. The viewing distance is 50 cm.

Procedure. Each threshold measurement is based on a run of 40 letter-identification trials. The identification trial is performed by the human, either unconstrained or as dictated by the kind of composite: two steps, one step, or none. Each correct identification is rewarded with a short beep. The observer is asked to fixate a central white dot subtending 0.10° on the monitor. The observer initiates the run by clicking a mouse. When the human acts alone or as a combiner, 1,000 ms later the stimulus appears for 200 ms, followed by a blank screen for 250 ms, followed by a noise-free response screen containing all of the letters. The observer uses a mouse-controlled cursor to select a letter from the response screen. Any response automatically initiates the next trial, 1 s later. When the human acts as a detector, he performs 18 feature-detection trials for each letter-identification trial. On each detection trial, he reports the presence or absence of the Gabor by key press. There is no detection-specific feedback; the only feedback is the identification reward at the end of the identification trial, i.e., after the 18th detection trial. The feedback indicates whether the human and ideal together chose the correct letter.

QUEST. The QUEST sequential estimation procedure provides threshold estimates over the course of learning (6). The QUEST procedure estimates from already-known information regarding both the task and observer (assumed stationary), as well as from the observer's performance throughout the run, to provide a maximum posterior probability estimate of threshold contrast, the signal contrast (ratio of luminance increment to background luminance) at which the observer correctly identifies the signal at criterion performance (75% correct). After each trial, the QUEST procedure calculates a threshold estimate. We place each new trial at the current threshold estimate. In practice, if the observer correctly identifies the signal, the next trial presents a lower contrast. If he incorrectly identifies the signal, the next trial presents a higher contrast. QUEST is initialized at the beginning of each run with log threshold estimate -1 ± 2 (\pm SD), β 3.5, lapse rate 0.01, and guess rate 0.125, and is updated after each identification trial.

Calculating the Slope of Learning. For each unconstrained or composite observer, for each participant, we fit a line to the data (log threshold contrast as a function of log trial) by linear least-squares regression. Extrapolating any of these rising lines makes the impossible prediction that the human will eventually beat the ideal. In fact, improvement must saturate eventually, after huge amounts of practice. Even so, Pelli et al. (7) found good straight-line fits to letter-learning data out to 50,000 trials. The ideal does not learn; it is unaffected by practice, so we display best-fit horizontal lines for I and II in Fig. 2.

SI Methods for Table 1

Here we provide the methods used to estimate the log-log slope of learning from the 13 studies presented in Table 1, top to bottom.

This Study, Composite Observer HI. The slope, -0.03 , is the average across all participants and is reported in the main text.

Pelli et al. (8), Familiar Letters. Experiment 3.4 of Pelli et al. (ref. 8, p. 4,658) measured improvement in threshold contrast for the identification of a letter. Participant RA performed 2,000 trials of the identification task using familiar letters. His efficiency increased from 6% (at 40 trials) to 7.3% (at 2,000 trials). We fit a straight line, in log coordinates, to these two points using linear least-squares regression; its slope was 0.050. Because efficiency is inversely proportional to threshold contrast squared, the log-log slope of efficiency is $-2\times$ that of threshold contrast. Therefore, the log-log slope of contrast learning is $0.050/-2 = -0.0250$. Two other participants, AW and DM, performed $\sim 2,500$ trials (in blocks of 40) of an identification task using 2×3 checkerboard

patterns. Figure 10 of ref. 1 (p. 4659) shows the data. The vertical axis plots the efficiency estimated from each block. The data are fit with a straight line on log-log axes. The slope of efficiency learning is 0.076 for participant AW and 0.100 for participant DM, and so the slope of contrast learning is -0.038 and -0.050 , respectively. Thus, the average log-log slope of contrast learning across the three participants is -0.04 .

Furmanski et al. (9). Furmanski et al. (ref. 9, figure 2a, p. 574) show improvement in threshold contrast for the detection of a Gabor. The learning curve is the average across six participants and shows learning over the course of a month. The reported "normalized threshold" is proportional to threshold and does not affect our estimate of the slope. We fit a line, in log coordinates, to the 34 normalized thresholds reported in the figure; its slope is -0.06 .

This Study, Composite Observer IH. The slope, -0.11 , is the average across all participants and is reported in the main text.

Suchow and Pelli (10). The figure in Result III shows improvement in efficiency for the identification of an unfamiliar letter from the Armenian alphabet. Two participants in Suchow and Pelli (10), SAS and JWS, performed 3,000 trials of the identification task in blocks of 40 trials. The vertical axis plots the efficiency estimated from each block. The data are fit with a straight line on log-log axes. The log-log slope of efficiency learning is 0.21 for participant SAS and 0.21 for participant JWS. Thus, the average log-log slope of contrast learning is -0.11 .

Pelli et al. (8), Unfamiliar Letters. We used the same method described above. Pelli et al. (ref. 8, figure 10, p. 4659) also reports seven learning curves for participants identifying unfamiliar letters. Each curve includes between 1,500 and 5,000 trials of an identification task. Participants SE, JB, and AW identified 4×4 checkerboard patterns; participants DM and AW identified Devanagari letters; participant AW identified Hebrew letters; participant JF identified English letters. The average log-log slope of contrast learning was -0.11 .

Lu and Doshier (11). Lu and Doshier (ref. 11, figure 4a, p. 50) show improvement in threshold contrast for the identification of the orientation of a Gabor tilted $\pm 8^\circ$ from diagonal. This task required a fine discrimination, which was initially unfamiliar to the participants. Participants were tested at each of two criteria (70.7% and 79.3% correct) at each of eight levels of added noise (rms contrast ranging from 0 to 0.33). In the text, the authors report that at the highest level of external noise, threshold contrast improved from 0.72 (for sessions 1 and 2 of 10, coded as session 1.5) to 0.48 (for sessions 9 and 10, coded as session 9.5). We fit a line, in log coordinates, to these two points; its slope was -0.22 . The slope is the same, -0.22 , if the line is instead fit to the data from all sessions at the highest noise level, not just the first and last two. Lower noise levels produced more shallow slopes of learning. The average slope across all noise levels and criteria is -0.12 (ranging from -0.014 to -0.24).

This Study, Observer H. The slope, -0.16 , is the average across all participants and is reported in the main text.

Gold et al. (12), Noise Texture. Gold et al. (ref. 12, figure 3, p. 177) show improvement in efficiency for the identification of a noise texture. For each of the two participants, AMC and JMG, we fit a straight line, in log coordinates, to the points using linear least-squares regression. The average log-log slope of contrast learning was -0.32 .

Gold et al. (12), Face. Gold et al. (ref. 12, figure 3, p. 177) show improvement in efficiency for the identification of a face.

For each of the two participants, AMC and CGB, we fit a straight line, in log coordinates, to the points using linear least-squares regression. The average log-log slope of contrast learning was -0.39 .

Michel and Jacobs (13). Michel and Jacobs (ref. 13, figure 6, p. 9) show improvement in efficiency for discrimination of shapes in filtered noise. The authors defined efficiency as the ratio of the sensitivity index d' of human and ideal, which is similar, but not identical, to our definition as the ratio of threshold energies. For each of the three participants who showed evidence of learning (BVR, WHS, and RAW), we fit a straight line, in log coordinates, to the points using linear least-squares regression. The average log-log slope of contrast learning was -0.48 .

Fine and Jacobs (14). Fine and Jacobs (ref. 14, figure 6, p. 3217) show improvement in threshold contrast for the discrimination of a complex plaid pattern. The high spatial frequency component of the plaid was placed at a different contrast than the low spatial frequency component, and so for analysis we separately measured the slope using the contrast of each component, and then averaged the slopes together to produce the final estimate. The across-participant average threshold contrast for sessions 1 and 2 (coded as session 1.5) was 0.081 and 0.27 for the low and high spatial frequency components, respectively. After the final sessions, 7 and 8 (coded at session 7.5), thresholds dropped to 0.024 and 0.078, respectively. We fit a straight line, in log coordinates, to the points using linear least-squares regression. The average log-log slope was -0.78 .

1. Watson AB, Robson JG (1981) Discrimination at threshold: Labelled detectors in human vision. *Vision Res* 21(7):1115–1122.
2. Kim J, Wilson HR (1993) Dependence of plaid motion coherence on component grating directions. *Vision Res* 33(17):2479–2489.
3. Kim J, Wilson HR (1993) Dependence of plaid motion coherence on component grating directions. *Vision Res* 33(17):2479–2489.
4. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.
5. Pelli DG, Zhang L (1991) Accurate control of contrast on microcomputer displays. *Vision Res* 31(7-8):1337–1350.
6. Watson AB, Pelli DG (1983) QUEST: A Bayesian adaptive psychometric method. *Percept Psychophys* 33(2):113–120.
7. Pelli DG, Burns CW, Farell B, Moore-Page DC (2006) Feature detection and letter identification. *Vision Res* 46(28):4646–4674.
8. Pelli DG, Burns CW, Farell B, Moore-Page DC (2006) Feature detection and letter identification. *Vision Res* 46(28):4646–4674.
9. Furmanski CS, Schluppeck D, Engel SA (2004) Learning strengthens the response of primary visual cortex to simple patterns. *Curr Biol* 14(7):573–578.
10. Suchow JW, Pelli DG (2005) Learning to identify letters: Generalization in high-level perceptual learning. *J Vis* 5(8):712, (abstr).
11. Lu ZL, Doshier BA (2004) Perceptual learning retunes the perceptual template in foveal orientation identification. *J Vis* 4(1):44–56.
12. Gold J, Bennett PJ, Sekuler AB (1999) Signal but not noise changes with perceptual learning. *Nature* 402(6758):176–178.
13. Michel MM, Jacobs RA (2008) Learning optimal integration of arbitrary features in a perceptual discrimination task. *J Vis* 8(2):, 3.1–16.
14. Fine I, Jacobs RA (2000) Perceptual learning for a pattern discrimination task. *Vision Res* 40(23):3209–3230.

