includes gravity, there are other interesting consequences of this line of argument. One of the most striking is the following. In QCD, when the temperature reaches sufficiently high values (above 10^{12} K), a phase transition occurs and quarks and gluons are no longer confined — instead, a 'soup' of free particles is formed, called quark–gluon plasma. In the five-dimensional theory, this transition also corresponds to the formation of a black hole in the interior. Our knowledge of black holes can then tell us something about quark–gluon plasma. In addition, the QCD–string theory provides a simple explanation for an interesting feature of black holes — the Bekenstein–Hawking entropy. This entropy, a measure of the number of possible quantum microstates, arises from the thermodynamic properties of a black hole (which are also at the root of Hawking radiation). Counting these microstates to work out the entropy has proved a major challenge in the-

ories of quantum gravity. But in the five-dimensional theory, the black-hole entropy becomes just the entropy of the plasma of quarks and gluons.

There is an intimate connection between the physics of strong interactions and both string theory and quantum gravity. Hopefully, in the next few years a string-theory description for real-world QCD will emerge, making it possible to perform computations in a relatively simple way. And perhaps, beyond that, we might even arrive at a QCD-like theory that can describe gravity. ∎

Juan Maldacena is in the School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey 08540, USA.
e-mail: malda@ias.edu

1. Polchinski, J. & Strassler, M. *J. High Energy Phys.* 05(2003)012.
2. 't Hooft, G. *Nucl. Phys. B* **72**, 461–473 (1974).
3. Maldacena, J. *Adv. Theor. Math. Phys.* **2**, 231–252 (1998).
4. Gubser, S., Klebanov, I. & Polyakov, A. *Phys. Lett. B* **428**, 105–114 (1998).
5. Witten, E. *Adv. Theor. Math. Phys.* **2**, 253–291 (1998).

Cognitive neuroscience

Practice doesn't make perfect

Wilson Geisler and Richard Murray

It may seem counterintuitive, but we are not very efficient at recognizing even the most common words. This finding suggests strict limits on how flexible we are in learning to recognize new patterns.

If there is any truth to the adage that 'practice makes perfect', we should certainly be near perfect at recognizing common words. As young children, we are taught first to recognize letters, and shortly thereafter words, and then for the rest of our lives we practise word recognition every day — the average literate person has read, at a conservative estimate, a hundred million words by the age of 25, thereby practising the recognition of each common word many hundreds of thousands of times. Nonetheless, on page 752 of this issue, Pelli, Farell and Moore[1] show that word recognition is surprisingly inefficient relative to letter recognition. This finding, which is surely not unique to words, implies that the neural learning mechanisms that are involved in pattern and object recognition are severely limited in their capabilities.

To measure the efficiency of word recognition, Pelli *et al.*[1] used the powerful tools and logic of ideal-observer analysis[2–6]. At the heart of this method is the concept of the 'ideal observer' — a theoretical device that achieves the best-possible performance in a perceptual or cognitive task, given the available information and constraints. Perceptual and cognitive tasks generally involve processing noisy physical or neural signals under conditions of uncertainty, and so ideal observers are typically derived using the concepts and methods of Bayesian statistics.

Pelli and colleagues' study elegantly demonstrates the value of this type of analysis. First, derivation of the ideal observer forces one to rigorously specify the task at hand and to determine at least one specific mechanism for achieving optimal performance. Here, the task was to identify words or letters embedded in spatial white noise (top of Fig. 1). In each trial, a word was randomly picked from a set of possible words, given a particular contrast, and then added to a random sample of noise; the contrast determined how well the word stood out against the noisy background.

One type of mechanism that achieves optimal word recognition in this task is the 'template matcher' (Fig. 1, left path). This mechanism stores an exact copy (a template) of each target (such as a word) that could appear in a stimulus. To identify a particular target, the template matcher calculates how well each point on the target corresponds with each point on each stored template, and identifies the target as the template that gives the highest correlation[7]. (If the various targets do not occur with equal probability, then each correlation is adjusted according to the target's probability.) Many neurons that are found in the first stages of the visual pathway can be modelled, to a first approximation, as template matchers.

The second benefit of ideal-observer analysis is that 'ideal' performance provides
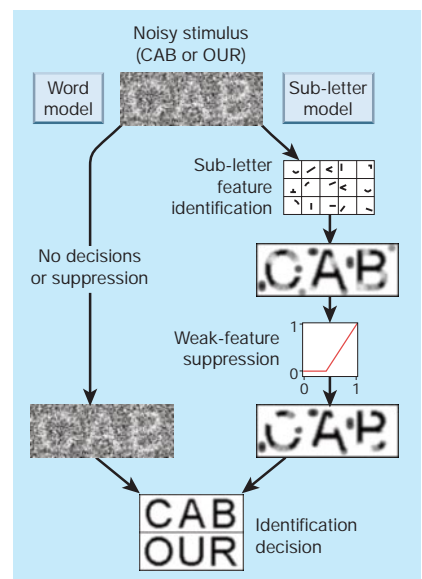

Figure 1 **Why are we so inefficient at recognizing words? An efficient word-recognition mechanism, such as a template matcher (left pathway), might integrate information over the entire stimulus. As more and more letters are added to a word, the template matcher can identify the word at lower and lower contrasts. On the other hand, a mechanism that suppresses weak neural responses to letters or parts of letters (sub-letter features; right pathway) sets a lower limit on the contrast at which words can be identified: below the suppression threshold, the stimulus is simply not visible. This threshold is represented in the graph by the point at which the curve relating input (*x*-axis) to output (*y*-axis) begins to increase. Consequently, the addition of more and more letters does not mean that words can be identified at lower and lower contrasts. Pelli *et al.*[1] have found that our ability to recognize words decreases rapidly as word length increases, implying that the brain does not work as a 'template matcher'. They argue that instead it functions more like the right pathway.**

an appropriate benchmark against which to compare human performance. In Pelli and colleagues' work[1], it makes it possible to rigorously answer the question of whether practice makes perfect, by providing a precise description of 'perfect' performance. The performance of a human observer relative to the ideal is called the efficiency. In this case, the efficiency equals the squared contrast of the target at which it can be identified by the ideal observer with some given level of accuracy (for instance, 70% correct), divided by the squared contrast required for the human observer to identify it with the same level of accuracy. Pelli and colleagues' main finding is that human observers' efficiency at recognizing words declines drastically as word length increases. Specifically, if a person's efficiency at recognizing single letters is F, then their efficiency at recognizing

n-letter words is F/n. So, for example, common five-letter words are identified with only 1/5 the efficiency of single letters. This relatively poor performance holds for even the most common three-letter words, which often occur many times in a paragraph.

Why are people so inefficient at recognizing even common words? The ideal observer can also be helpful in addressing such questions, by providing a starting point for formulating possible models of human (sub-optimal) performance. We know that, with a little practice, humans can become extremely efficient at recognizing some simple patterns — such as spatially localized bars, edges and spots (see, for example, ref. 4). This is probably possible because the visual cortex in the primate brain contains neurons that act like template matchers for these simple patterns. But the $1/n$ efficiency relationship reported by Pelli *et al.* implies that the brain cannot learn optimal templates for whole words. Rather, their results are quantitatively consistent with a model in which the brain has efficient templates only for parts of words — letters, or even features of letters — and in which only the strong responses of the neurons encoding these parts are passed on to word-identification processes (Fig. 1, right path). The hypothesis that familiar complex objects are initially encoded in smaller parts is consistent with recent psychophysical studies of the identification of objects in noise[8].

Although Pelli and colleagues' results[1] show that word recognition is relatively inefficient, they do not rule out the possibility of specific neural circuits for the recognition of common words or other familiar objects. For example, there could be individual neurons whose receptive fields are precisely tuned to such objects. What the results do imply is that any neural circuits for recognizing complex objects such as words (or letters) cannot integrate neuronal responses as efficiently as can, say, the circuits underlying the receptive fields of neurons in the primary visual cortex (an early stage of the visual pathway), which encode simple patterns such as bars, edges and spots.

Pelli and colleagues suggest that inefficient integration occurs because weak responses to features (or letters) are suppressed ('squelched'). Moreover, they suggest that the brain makes discrete (categorical) decisions about which features (or letters) are present before deciding which word is present. Although this is plausible, there are alternative explanations that are similar in spirit but differ in detail. For example, it might not be necessary to invoke the suppression of weak responses to features. The $1/n$ efficiency relationship could also be a result of increasing noise (neural noise, for instance) or uncertainty at lower target contrasts. The essential thing is that the signal-to-noise ratios of the neuronal responses

that are passed on to the word-identification process degrade rapidly as the target's contrast is lowered. Similarly, discrete decisions about features or letters are not required; simulations that we have carried out suggest that the predictions are almost the same if graded responses are used in word identification, as long as the graded responses deteriorate at low contrasts.

Regardless of the specific explanation, Pelli and colleagues' results[1] clearly show that, even with extensive practice over many years, perceptual learning mechanisms are unable to achieve efficient recognition of arbitrary objects. This finding is particularly revealing given that the visual system is very efficient at recognizing certain simple patterns such as bars and edges. It seems that Pelli *et al.* have uncovered, with the help

of ideal-observer analysis, a fundamental limit on the neural mechanisms of learning and plasticity.   ■

Wilson Geisler and Richard Murray are at the Center for Perceptual Systems and the Department of Psychology, University of Texas, Austin, Texas 78712, USA.
e-mail: geisler@psy.utexas.edu

1. Pelli, D. G., Farell, B. & Moore, D. C. *Nature* **423**, 752–756 (2003).
2. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
3. Barlow, H. B. *Vision Res.* **18**, 637–650 (1978).
4. Burgess, A. E., Wagner, R. F., Jennings, R. J. & Barlow, H. B. *Science* **214**, 93–94 (1981).
5. Geisler, W. S. *Psychol. Rev.* **96**, 267–314 (1989).
6. Geisler, W. S. & Kersten, D. *Nature Neurosci.* **5**, 508–510 (2002).
7. Peterson, W. W., Birdsall, T. G. & Fox, W. C. *Trans. Inst. Radio Engineers Prof. Group Information Theory* **4**, 171–212 (1954).
8. Gold, J. M., Murray, R. F., Bennett, P. J. & Sekuler, A. B. *Curr. Biol.* **10**, 663–666 (2000).

Astrophysics

Superficial resonance

Frits Paerels

Neutron stars are the most poorly understood stellar objects in the Universe. But observations of X-rays emitted from one neutron star have now revealed a clue to the nature of its surface and composition.

Neutron stars are the remnants of massive stars. The density of their cores exceeds that of atomic nuclei, and they might contain exotic states of matter that are not found anywhere else in nature[1]. For decades astronomers have tried to 'see' the surfaces of such stars, in an effort to determine their physical properties. Space-borne telescopes have brought improved X-ray-imaging capabilities, but the first observations of neutron stars with these instruments were rather disappointing, revealing that the X-ray emission from most isolated neutron stars is remarkably featureless. But now observations of one neutron star have finally given us a glimpse of the fundamental properties of these objects: on page 725 of this issue, Bignami *et al.*[2] describe the first physical clue to the nature of this star's X-ray emission.

When massive stars — stars more than eight times heavier than the Sun — run out of thermonuclear fuel, their cores collapse. The resulting configuration typically has a mass about one-and-a half times that of the Sun, but a radius of only 10 km: the density of the material exceeds the nuclear density, and degeneracy (the quantum-mechanical restriction on the occupation of energy levels by particles) provides the support against the gravity of the star's own weight. According to simple astrophysical arguments, a young neutron star is expected to have a magnetic field of the order of 10^{12} gauss at its surface (in comparison, Earth's field is about 0.5 gauss). It should also spin rapidly, appearing as a 'pulsar'.

The first neutron stars were discovered through their emission at radio wavelengths. Radio emission is produced in the envelope of magnetic field that surrounds a pulsar (through processes that are still poorly understood), and although it carries information about the star's magnetic field and spin period, it tells us nothing about the internal properties of the neutron star itself. If, instead, we could observe emission from the star's surface, it would tell us about the surface composition, which in turn would reveal more about how a neutron star forms, and the processes that affect the chemical composition of its atmosphere.

To 'view' the surface of hot young neutron stars requires observations of their X-ray, rather than radio, emission: the emission peaks in the 'soft' X-ray band of photon energies (0.1–1 kiloelectronvolts, or keV). But even in the X-ray band, isolated hot neutron stars are faint. The 1999 launches of NASA's Chandra X-ray Observatory and the European Space Agency's XMM-Newton observatory were eagerly awaited, but the initial observations of a handful of hot neutron stars were universally disappointing. The spectra of radiation from the stars did show the expected characteristic of thermal emission from their surfaces: a gently curving, black-body-like energy distribution. But detailed scrutiny of the X-ray spectrum failed to reveal any structure; the shape of the spectrum in all cases was very close to that of a black body, which by its very nature

....

The remarkable inefficiency of word recognition

Denis G. Pelli*, **Bart Farell**† & **Deborah C. Moore**†

* *Psychology and Neural Science, New York University, New York, New York 10003, USA*
† *Institute for Sensory Research, Syracuse University, Syracuse, New York 13244-5290, USA*

....

Do we recognize common objects by parts, or as wholes? Holistic recognition would be efficient, yet people detect a grating of light and dark stripes by parts. Thus efficiency falls as the number of stripes increases, in inverse proportion, as explained by probability summation among independent feature detectors[1]. It is inefficient to detect correlated components independently. But gratings are uncommon artificial stimuli that may fail to tap the full power of visual object recognition. Familiar objects become special as people become expert at judging them[2,3], possibly because the processing becomes more holistic. Letters and words were designed to be easily recognized, and, through a lifetime of reading, our visual system presumably has adapted to do this as well as it possibly can. Here we show that in identifying familiar English words, even the five most common three-letter words, observers have the handicap predicted by recognition by parts: a word is unreadable unless its letters are separately identifiable. Efficiency is inversely proportional to word length, independent of how many possible words (5, 26 or thousands) the test word is drawn from. Human performance never exceeds that attainable by strictly letter- or feature-based models. Thus, everything seen is a pattern of features. Despite our virtuosity at recognizing patterns and our expertise from reading a billion letters, we never learn to see a word as a feature; our efficiency is limited by the bottleneck of having to rigorously and independently detect simple features.

The role of components in object recognition is still mysterious[4–7]. For most objects we don't even know what the components are, but we do know that words are made of letters. Is a familiar word recognized as an image, or as a combination of individually recognized letters? Despite a century of careful study[8–16], it has never been noted that these two alternatives predict very different thresholds for identifying words. Once we have defined a few terms, you can test the predictions with your own eyes.

The strength of visual signals is traditionally specified by 'contrast', which here is the ratio of the luminance increment of the letter or word to the background luminance. However, for ideal-observer analysis it is helpful to specify 'contrast energy', which for a letter or a word is the product of squared contrast and 'ink' area. In general, contrast energy is the integral of the squared signal contrast over the extent of the stimulus. Energy matters: mathematical work on radar in the 1950s proved that the energy of a known signal completely determines its detectability in white noise[17]. 'Threshold', the border between seeing and not seeing, is defined here as the contrast or energy required by the observer to correctly identify the letter or word 64% of the time.

In Fig. 1a, the two lines of faint text have the same overall contrast energy, but differ in the way that the energy is distributed. To optimize recognition by parts, the first line gives each letter the same energy. To optimize recognition as wholes, the second line gives each word the same energy. Recognition by parts predicts that the longer words in the second line will be illegible, as you see, because there isn't enough energy per letter. Recognition as wholes predicts that all words in the second line will be equally legible, contrary to what you see. Figure 1b shows the predictions for your word threshold on blank and noisy backgrounds. The same analysis applies to both backgrounds because the observer effectively adds his or her intrinsic visual noise to the display[18].

Our recognition-as-wholes predictions are based on the ideal observer, which, given the stimulus and its statistics, achieves the best possible expected performance by choosing the most probable letter or word[17,18]. We can assess human performance on an absolute scale by defining 'efficiency' as the ratio of the ideal's threshold energy to the human observer's: the fraction of the energy used by
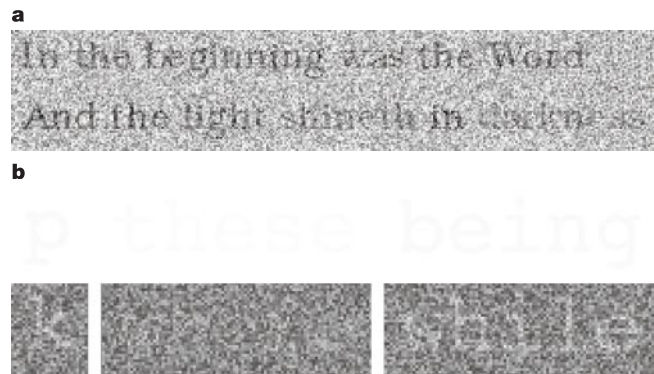

Figure 1 By letter or by word? **a**, Both lines of the quotation have the same total contrast energy. In the first, the energy is divided equally among the letters. In the second, the energy is divided equally among the words, regardless of length. In principle, at a given noise level a pattern's detectability depends only on its energy, but in the second quotation the short words pop out and the long words disappear. This word-length effect shows that human readers cannot efficiently integrate the energy across a whole word. (The quotation reads 'In the beginning was the Word … And the light shineth in darkness'.) **b**, Two predictions for the word threshold. The left column shows a letter at threshold on a uniform white background (top row) and on a noisy background (bottom row). They may take a minute to appear. The middle column shows a 5-letter word at the same energy ($1/\sqrt{5}$ the contrast), which is the threshold predicted by recognition as wholes, and the right column shows a word at 5 times the energy (the same contrast), which is the threshold predicted by recognition by parts. The words in the middle column would be identifiable if you could see words as efficiently as you see letters. (The first line reads 'p', 'these', 'being', and the second reads 'k', '?????', 'while', where the identity of the middle word remains undisclosed.) Note: The faint lettering on a white background is at the limits of what can be rendered on the printed page. The PDF of this letter prints the figure successfully on most modern printers, especially colour printers. If in doubt, readers are urged to refer to a more robust version of Fig. 1b available as Supplementary Information, which accommodates variations in printers' rendering and readers' sensitivity.

the observer that is ideally required to account for the observer's measured performance[18,19]. This strips away the intrinsic difficulty of the task, exposing a pure measure of human ability.

Our argument begins with our finding that human efficiency for word identification is inversely proportional to word length, independent of the number of possible words, as predicted by recognition by parts with suppression of weak signals, 'squelching'. Then we show that, notwithstanding the well known 'word superiority effect', human word identification never exceeds the accuracy attainable by strictly letter-based models. Related work indicates that letters are made of features. Finally, we conclude that visual recognition of even the most familiar objects is severely restricted by a first stage of independent feature detectors that squelch and each integrate no more than a letter, and probably much less.

The word-length effect, for words 2–16 letters long, is shown in Fig. 2 for two observers. Figure 2a shows that their threshold energy is proportional to word length, whereas the ideal observer's
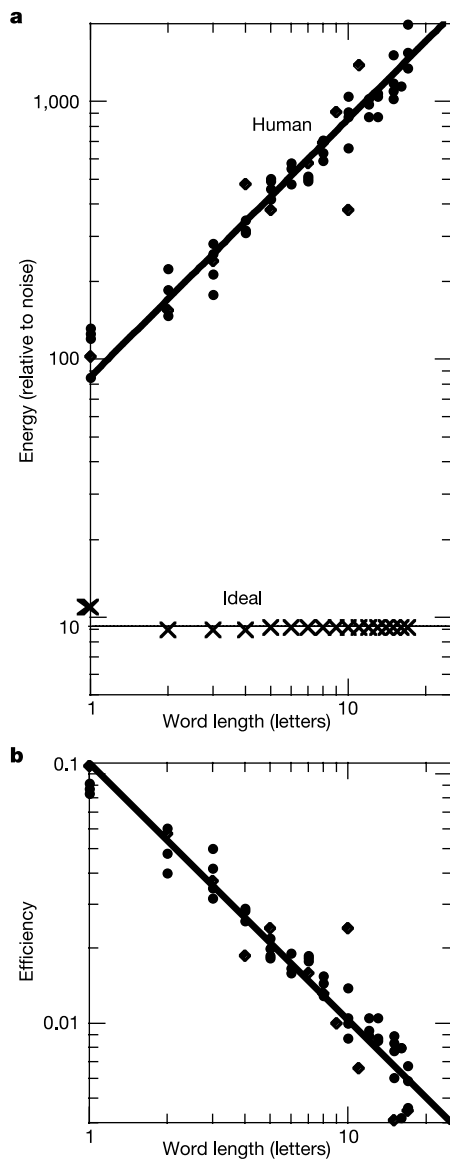

a

Figure 2 The effect of word length. **a**, Threshold energy for identifying one of 26 words, as a function of word length. The data at length 1 are for single letters. The ideal observer's thresholds (crosses) lie near the zero-slope line. The human observers' thresholds (UP, circles; JG, diamonds) lie near the unit-slope line. **b**, Efficiency (the ratio of ideal and human thresholds) derived from **a** as a function of word length. The points are close to the line with −1 slope: efficiency is inversely proportional to word length.

threshold is practically independent of word length. Figure 2b shows that efficiency, the ratio of ideal and human thresholds, is inversely proportional to word length. Efficiency for n-letter words is $1/n$ that for single letters. This result is not surprising for longer words, as observers take in only a modest number of letters in a glimpse, perhaps 4.5 (ref. 20), which predicts the reciprocal drop in efficiency for longer words, as the ideal observer uses all the letters. What is surprising and important about Fig. 2 is that the same slope extends left to the shortest words, and even single letters. Thus, the required energy per letter is independent of word length.

This is not merely a consequence of just size or contrast. Although a word is bigger than a letter, and words and letters have the same threshold contrast, the human limitation demonstrated in Figs 1 and 2 is neither an inefficiency for all large objects nor a fixed threshold contrast that all objects must exceed to be seen. Increasing letter size fivefold, to match the width of a word, reduces the letter's threshold contrast, in noise, fourfold[21]. Because seeing the large letter requires only a fraction of the contrast required to see a similar-width word, any explanation of the poor visibility of words must penetrate past their size to consider their internal structure. Our results indicate that, rather than directly recognizing complex familiar objects, such as words, our visual system detects smaller components—letters or perhaps features of letters—and only then recognizes the object specified by these components. Nothing is seen unless its components are detected[22,23].

One might ask whether the human observer could reasonably be expected to confine his or her word search to the 26 words used in the test, as the ideal does. Our design minimized this concern by using the 26 most common n-letter words, and displaying them as the response alternatives in every trial. Identifying letters independently (recognition by parts) is inefficient because the letters in a word are correlated. That's because the number of n-letter words, 26 in our experiment or thousands in real life, is a tiny fraction of the number of possible strings, 26^n. Still, one might worry that this human inefficiency is an artefact of using only 26 words. We addressed these concerns by measuring human and ideal thresholds for identification of one of 2,213 five-letter words (the most common 2,213), presenting each word at the same relative frequency as it appears in print[24], and found the same efficiency (approximately 4%) as for the 26 most common. (These efficiencies, for experienced observers with Courier font, are slightly higher than in Fig. 2, which is for new observers with Bookman font. The 100,000-trial experience, in various conditions, and the Courier font both contribute to the ×1.5 higher efficiency[21].) Finding the same efficiency means that human and ideal thresholds are affected equally by word frequency. Thinking that perhaps only extremely common words enjoy recognition as elementary visual patterns[25], we measured human and ideal thresholds for identifying the five most common 3-letter words (the, and, was, for, his). The least frequent of these words (his) is encountered 100 times in an hour of reading, nearly 400,000 times in a decade of reading an hour a day. Even so, the five most common 3-letter words yield practically the same efficiency as the 26 most common (4.8% compared to 4.5%) and reciprocity holds: the efficiency for identifying the five most common 3-letter words is 1/3 that for single letters: 4.8%/15%. Changes in intrinsic task difficulty—5, 26 or 2,213 alternatives—invalidate comparison of raw thresholds unless modelling assumptions are made, but efficiencies are always directly comparable.

The ideal observer chooses the best-matching template, using one template for each possible word. The root mean square (r.m.s.) difference between the stimulus and the template is a measure of the likelihood that the stimulus is the template plus noise. To choose the most probable word the ideal observer weighs each word's likelihood by its frequency. Template matching integrates energy efficiently over the extent of the template. If humans did template matching, with accurate templates for words of every length, then the slope in Fig. 2b would be zero, not −1. Figure 2 indicates the

753

absence of templates of more than one letter, as the energy beyond one letter doesn't reduce the required energy for the first letter.

The brain is well equipped to do template matching. A typical neuron sums over 10,000 synapses, each with a different gain. Any neuron that integrates over part of the visual field computes the likelihood of the presence of a signal matching the neuron's sensitivity profile. Thus, neurons with very simple receptive fields have been called "fly detectors"[26]. One can speculate that there might be neurons (perhaps in brain area IT) that linearly integrate over more complex receptive fields that match a face, letter or word, but such neurons would allow the observer to attain much higher efficiencies than found here, placing their existence in doubt.

In principle, identifying a letter independently requires the same energy for that letter as would be required to identify the whole word (if there are, as in Fig. 2, the same number of possible letters and words). The fact that a word is unreadable by our observers unless its letters are separately identifiable is evidence for recognition by components; that is, identification of the word is mediated by independent detection of components that are a letter or less. We define 'features' as image components that are detected independently, unaffected by the presence of other features. Independent feature detection is a bottleneck, especially when feature thresholds are high; complex objects will be visible only when the energy per feature reaches threshold.

Our data indicate that there are no multi-letter features. Efficiency for letters is independent of age after ten, only weakly dependent on size and alphabet (English, Armenian, Devanagari and Hebrew), and inversely proportional to complexity[21]. 'Complexity' is a scale-invariant physical measure: perimeter squared over 'ink' area[21]. The number of features in a letter may be proportional to its complexity. It seems that in identifying letters, observers use no feature more complex than the average letter in the simplest alphabet tested[21], about one-third the complexity of the Bookman and Courier fonts used here. As letters and words are designed to be legible, and the reader's visual system presumably has adapted itself to them as much as it can, we conclude that the feature bottleneck is unavoidable. Objects are recognized by means of independent detection of their component features, which are much simpler (less complex) than a single Bookman letter.

The efficiency result—the reciprocal relation between efficiency and number of components—is secure and assumption-free, but how general is it? We extended it from obscure gratings[1] to common words. It also applies to letters if we suppose that complexity is proportional to the number of features. But these stimuli were all at threshold. Threshold stimuli are directly relevant to real-life reading of highway signs, which are usually read at great distance as soon as they become readable. Might ordinary reading, at high contrast, be mediated by different mechanisms? That seems unlikely. Critical-band masking studies have characterized the channels (feature detectors) that mediate letter identification at threshold[27]. Experiments at high supra-threshold contrasts, measuring the effect of noise on reading rate, reveal the same channel tuning[28].

We usually see things quite reliably or not at all, with a fairly abrupt transition between the two. Indeed, the 'psychometric function', the probability of seeing, rises much more steeply as a function of contrast than predicted by theory of signal detectability for an exactly known signal[17]. The steep psychometric function means that weak signals are suppressed. Engineers call this 'squelching': in better walkie talkies, a nonlinear analog circuit turns down the volume when the signal is weak relative to the background noise, to cut out the hiss when no one is speaking. In the same way, human vision squelches features, allowing them to pass only if they are well above the noise. We call this 'detecting rigorously'. It achieves reliability at the expense of efficiency. The human visual system has a vast number of feature detectors, each of which can raise a false alarm, mistaking noise for signal. Squelching blocks the intrusion of countless false features that would besiege us if weak features were

not suppressed. However, it impairs our ability to see more complex objects, like words. In the limiting case, squelchers pass all signals above a certain threshold contrast, and none below. Whether the squelch is gradual or abrupt, it is, in effect, deciding whether or not a signal is present and acting on that decision.

Discussions of letter and word recognition typically suppose various stages in the recognition process: identification of features, then letters, then the word[14,29]. One approach to modelling how an observer performs a task is to specify just the first stage of processing, leaving subsequent stages unspecified. However, such models must avoid the common mistake of specifying a transformation that preserves all task-relevant information. Such a transformation can always be undone by the subsequent unspecified stages, so it does not constrain performance, making the model psychophysically inconsequential and untestable. Proposing a first stage constrains performance only to the extent that the stage discards task-relevant information. In that spirit, when we assert that a part of a model, or the brain, 'makes a decision', we mean not only that it passes that decision on, but also that it discards (fails to pass on) the information the decision is based on. Thus, we consider the idea that all word and letter identifications are strictly letter-based; that is, they depend on the visual stimulus solely through a stage that identifies each letter independently and passes on only that identification, discarding the rest of the stimulus information. We leave the later stages, after letter identification, unspecified. Each decision is a bottleneck.

Selfridge's Pandemonium Model[29] cascades row after row of 'demons' that compute the likelihood of features, then letters, and then words, achieving perfect efficiency by not discarding any choice-relevant information until the final shouting match. It is an ideal observer. Similarly, the Interactive Activation Model[14] computes each possible letter's likelihood at each letter position, and postpones the information-discarding shouting match until the final response-selection stage. Presumably, given optimal weights and extended to report whole words, it would be nearly ideal, so it too would have similar energy thresholds for words and letters, unlike what we report here for human observers.

It is a simple matter to calculate how accurately a word can be identified by a strictly letter-based observer. The observer's first stage identifies the letters independently. These identifications are tentative. To perform optimally, the observer, having the string of tentatively identified letters, uses a historical table of his or her single-letter confusion probabilities (identifying one letter as
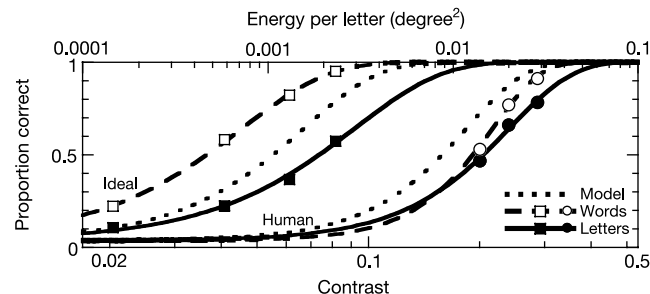

Figure 3 Proportion correct in identifying a letter (filled symbols) or a 5-letter word (open symbols) in noise as a function of contrast. Average energy per letter is indicated on the upper x axis. This is for one human, WT (3,000 trials per point), and the ideal observer (100,000 trials per point). Note that the human's contrast thresholds (64% correct) for letters and words are similar, whereas the ideal observer's are very different. For each observer, the dotted line is the best possible strictly letter-based word identification performance, given the observer's measured single-letter confusion probabilities (of identifying each letter as another) at that contrast. (A second human observer, DM, gave very similar results, not shown.) All curves are maximum-likelihood Weibull fits ($\gamma = 1/26$)[21].

another) at that contrast to choose the most probable word from the list of alternatives. (The ideal, described earlier, computes each possible word's likelihood based on the stimulus. Our letter-based two-stage observer is not ideal, and must settle for likelihood based on the independent letter identifications, rather than the stimulus itself.)

The performance of this best-possible letter-based observer is plotted as the dotted curves in Fig. 3. The left curve is based on the ideal's single-letter confusion probabilities at each contrast, and the right curve is based on the human's. Each dotted curve is the best possible accuracy for strictly letter-based word identification by that observer, given the observer's measured letter performance. Note that the human's word performance (right dashed line) never exceeds this letter-based bound, whereas the ideal's word performance (left dashed line) far exceeds it. Thus, despite reading for decades, a hundred million words[21], our observers identify even the most common words with an accuracy attainable through strictly letter-based identification.

Humans squelch features. The ideal observer does not. Nor does the two-stage model based on the ideal letter identifier. It is inefficient to detect correlated components independently, and this is exacerbated by squelching. We can calculate efficiency from the threshold contrasts of the curves in Fig. 3 for lengths 1 (letter) and 5 (word). Efficiency of the model, which doesn't squelch, drops as words grow longer. Efficiency of the human, who does squelch, drops more, as the reciprocal of word length.

At first sight, our conclusion may seem incompatible with the hundred-year-old 'word superiority effect', whereby a letter within a word is recognized better than a letter presented in isolation or in a scrambled word[8–16]. The word context improves letter identification even when it provides no information that can distinguish between the response alternatives. For example, coin versus join is easier than c versus j. Thus, in Fig. 4, the proportion correct for words (dashed line) is higher than that for letters (solid line). But is the observed context effect big enough to prove that the internal letter identifications are not independent of each other? No, it's too small. The human performance plotted in Fig. 4 is consistent with strictly letter-based word identification. The dotted line shows how well the observer would perform by first identifying the test letter as a ... z, and then choosing the more likely of the two response alternatives (for example, 'c' or 'j') based on her own single-letter confusion probabilities. This is the best possible strategy for an observer who is strictly letter-based. This letter-based upper bound applies to both letter and word conditions, and in fact is above both. The worse human performance shows that the observer is not using the optimal strategy, quite possibly because she doesn't know her own

single-letter confusion probabilities and thus fails to choose the most probable letter or word (c or j) given her tentative internal letter identification (a ... z). The strictly letter-based upper bound is also above the word advantage found in other studies, which find the proportion correct for letter identification to be 0.05–0.15 greater in a word context[10–16]. The word context (dashed line) improves performance, bringing it closer although still not exceeding the (dotted) upper bound. The slightly higher human performance in a word context (the word superiority effect) suggests that observers more accurately incorporate their historical confusion probabilities when reading words than when identifying letters, which is consistent with the observation that observers perform better if they attend to the entire word rather than just to the target letter[12]. Perhaps years of fast reading have trained the second-stage word-recognition process to learn the observer's letter-confusion probabilities, to more efficiently map strings of tentatively identified letters to real words. Figure 4 shows that the 0.10 upward increase in accuracy corresponds to a factor of ×1.15 leftward reduction in threshold contrast (×1.3 in energy). Thus, there are not one but two effects, one small and one large. The word superiority effect increases efficiency by a mere factor of ×1.3. The word-length effect, described here, is big, reducing efficiency by the word length, ÷5 for a 5-letter word. Both effects are consistent with strictly letter-based word identification.

In evaluating whether recognition is by parts or as wholes, we took a 'best of breed' approach, considering the best possible performance of models that conform to simple assumptions about the observer's internal processes. Our analysis has focused on letters because they are obvious components of words. More generally, the same approach may be applied to critically test any conjecture that object recognition is mediated by independent decisions about a specified set of features or components. The components in objects are usually correlated, so object recognition based on independent decisions about components is inefficient, especially with squelching. Having to independently and rigorously detect the components makes efficiency inversely proportional to their number. □

Methods

Testing

In each trial the signal was one of 26 possible letters (a ... z), or 26 possible words. The word list for each length (2–16 letters), for example 'of, to, in, ..., pa' for 2-letter words or 'responsibilities, misunderstandings, characterization, ..., aristocratically' for 16-letter words, consisted of the 26 most common words of that length, excluding words that are normally capitalized (such as 'American'), words containing punctuation, and nearly identical words[24]. The signal—a letter or word—was briefly presented (200 ms) either on a blank 50 cd m^{-2} gamma-corrected screen (for human observers) or in gaussian noise (for both human and ideal observers). The blank screen and the noise had the same mean luminance (unlike the two rows of Fig. 1b). The task was to identify the signal by choosing one of the 26 signals, which were displayed for the human observer on an immediately-following response screen. (Increasing the viewing time, as in the printed demonstrations in Fig. 1, does not affect identification on the static noise background, but does aid identification on a blank background, because detection is then limited by the observer's intrinsic visual noise[18], which is dynamic. See also ref. 15.) The text was rendered at 29 point in an off-screen image using the uniformly spaced TrueType Courier font (Fig. 1b), except for Figs 1a and 2, for which we used the proportionally spaced PostScript Adobe Bookman font. Efficiency of letter identification is slightly higher using Courier than using Bookman[21]. In the noise conditions, independent zero-mean gaussian samples, with standard deviation equal to 25% of the mean luminance, were added to the pixels of the off-screen image. The off-screen image was then doubled in size horizontally and vertically by pixel replication, and copied to the screen. The screen displayed 31.3 pixels per degree at the 0.6 m viewing distance. The power spectral density of the noise was $N = 10^{-3.59}$ deg^2. The typographic x-height of the displayed text was 0.83 deg (Courier) or 0.89 deg (Bookman).

Modelling

The results of the simulations at three contrasts were fit with a Weibull psychometric function, which is displayed as the dotted curve. The 26 × 26 table of probabilities of each one-letter response to each one-letter signal is the observer's letter-confusion matrix. The simulation used 3,000-trial letter-confusion matrices measured for each observer at each of three contrasts in noise. (For the ideal they were 100,000-trial matrices at four contrasts.) To prevent bias due to correlated sampling errors between the confusion matrices used to simulate the first and second stages of the model, each stage used an
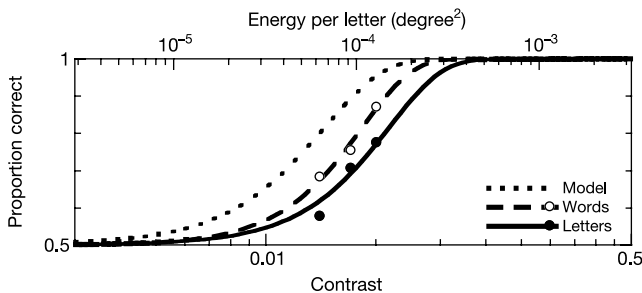

Figure 4 The word superiority effect. Proportion correct as a function of contrast and average energy per letter, for letters and words on a blank background. The dotted line represents the upper bound for strictly letter-based word identification, based on the observer's measured confusion probabilities at three contrasts. Unlike Fig. 3, only two response alternatives, differing by only one letter, were offered on each trial, for example, 'coin' versus 'join' or 'c ' versus 'j '. The word superiority is slight, no greater than can be accounted for by strictly letter-based identification.

755

independent 1,500- (or 50,000-) trial subset of the empirical letter-confusion counts. The first stage receives a letter i_1 or a word i_1, i_2, i_3, i_4, i_5 and independently emits a letter j_1 or a letter at each position j_1, j_2, j_3, j_4, j_5 with each letter probability specified by the 26×26 confusion matrix $P(j \mid i)$. The best-possible second stage chooses the most probable 5-letter word i_1, i_2, i_3, i_4, i_5 given the independent first-stage letter identifications j_1, j_2, j_3, j_4, j_5; that is, it maximizes the posterior probability

$$P(i_1, i_2, i_3, i_4, i_5 \mid j_1, j_2, j_3, j_4, j_5)$$

$$= \frac{P(i_1, i_2, i_3, i_4, i_5)P(j_1 \mid i_1)P(j_2 \mid i_2)P(j_3 \mid i_3)P(j_4 \mid i_4)P(j_5 \mid i_5)}{\sum_{i_1, i_2, i_3, i_4, i_5} P(i_1, i_2, i_3, i_4, i_5)P(j_1 \mid i_1)P(j_2 \mid i_2)P(j_3 \mid i_3)P(j_4 \mid i_4)P(j_5 \mid i_5)}$$

where $P(i_1, i_2, i_3, i_4, i_5)$ is the prior probability of the word i_1, i_2, i_3, i_4, i_5. Concerned that the resulting 1,500 (3,000/2) trials per confusion matrix might not be enough, we simulated the two-stage model based on 100, 1,000, 3,000, 33,000 and 100,000 trials of the ideal observer. This revealed that the proportion correct is robust, only a few per cent lower for simulations using confusion matrices based on 3,000/2 rather than on 100,000/2 trials. Thus, our dotted curves in Figs 3 and 4 slightly underestimate the best-possible strictly letter-based performance, making our conclusion slightly more secure: human performance never clears the bar.

Word superiority

Following Reicher's[10] elegant design, we began with a list of 288 pairs of 4-letter words that differed by only one letter within each pair. Equal numbers of words differed at each of the four letter positions. Our list was derived from that of ref. 30, replacing 18 obscure words (such as boll, lave, wile) by more common ones (ball, lake, mile). The observer was EG. In the word-identification task the observer was shown a low-contrast word, randomly selected from the list. Unlike our previous experiments, the response screen merely asked the observer to select between the correct word and its mate, which differed in only one letter position, as in 'coin' versus 'join'. The letter-identification task drew from the same word list, but left blank all but the differing letters in each pair ('c　　' versus 'j　　') in both the stimulus and response screens. The ideal observer performs identically on both tasks, because the non-differing letters are irrelevant to the choice, so the human observer's word superiority implies that here the human is identifying words slightly more efficiently than letters. Thus, the word superiority effect is aptly named, but, as explained in the text, is consistent with strictly letter-based word identification.

1. Robson, J. G. & Graham, N. Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Res.* **21,** 409–418 (1981).
2. Diamond, R. & Carey, S. Why faces are and are not special: an effect of expertise. *J. Exp. Psychol. Gen.* **115,** 107–117 (1986).
3. Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neurosci.* **3,** 191–197 (2000).
4. Treisman, A. & Schmidt, H. Illusory conjunctions in the perception of objects. *Cogn. Psychol.* **14,** 107–141 (1982).
5. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94,** 115–147 (1987).
6. Tarr, M. J. & Buelthoff, H. H. Image-based object recognition in man, monkey and machine. *Cognition* **67,** 1–20 (1998).
7. Pelli, D. G., Palomares, M. & Majaj, N. J. Crowding is unlike ordinary masking: distinguishing feature detection and integration. *J. Vision* [online] ⟨http://www.journalofvision.org/⟩ (in the press).
8. Cattell, J. M. The time taken up by cerebral operations. *Mind* **11,** 220–242 (1886).
9. Huey, E. B. *The Psychology and Pedagogy of Reading* (Macmillan, New York, 1908).
10. Reicher, G. M. Perceptual recognition as a function of the meaningfulness of stimulus material. *J. Exp. Psychol.* **81,** 275–280 (1969).
11. Wheeler, D. D. Processes in word recognition. *Cogn. Psychol.* **1,** 59–85 (1970).
12. Johnston, J. C. & McClelland, J. L. Perception of letters in words: seek not and ye shall find. *Science* **184,** 1192–1194 (1974).
13. Johnston, J. C. & McClelland, J. L. Experimental tests of a hierarchical model of word identification. *J. Verbal Learn. Verbal Behav.* **19,** 503–524 (1980).
14. McClelland, J. L. & Rumelhart, D. E. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychol. Rev.* **88,** 375–407 (1981).
15. Prinzmetal, W. & Silvers, B. The word without the tachistoscope. *Percept. Psychophys.* **55,** 296–312 (1994).
16. Jordan, T. R. & Bevan, K. M. Position-specific masking and the word–letter phenomenon: reexamining the evidence from the Reicher-Wheeler paradigm. *J. Exp. Psychol. Hum. Percept. Perform.* **22,** 1416–1433 (1996).
17. Peterson, W. W., Birdsall, T. G. & Fox, W. C. Theory of signal detectability. *IRE Trans. Inf. Theory* **4,** 171–212 (1954).
18. Pelli, D. G. & Farell, B. Why use noise? *J. Opt. Soc. Am. A* **16,** 647–653 (1999).
19. Tanner, W. P. Jr & Birdsall, T. G. Definitions of d' and η as psychophysical measures. *J. Acoust. Soc. Am.* **30,** 922–928 (1958).
20. Legge, G. E., Pelli, D. G., Rubin, G. S. & Schleske, M. M. Psychophysics of reading—I. Normal vision. *Vision Res.* **25,** 239–252 (1985).
21. Pelli, D. G., Burns, C. W., Farell, B. & Moore, D. C. Identifying letters. *Vision Res.* (in the press).
22. Campbell, F. W. & Robson, J. G. Application of Fourier analysis to the visibility of gratings. *J. Physiol. (Lond.)* **197,** 551–566 (1968).
23. Watson, A. B. & Robson, J. G. Discrimination at threshold: labelled detectors in human vision. *Vision Res.* **21,** 1115–1122 (1981).
24. Kucera, H. & Francis, W. N. *Computational Analysis of Present-Day American English* (Brown Univ. Press, Providence, 1967).
25. Hadley, J. A. & Healy, A. F. When are reading units larger than the letter? Refinement of the Unitization Reading Model. *J. Exp. Psychol. Learn. Mem. Cogn.* **17,** 1062–1073 (1991).
26. Barlow, H. B. Summation and inhibition in the frog's retina. *J. Physiol. (Lond.)* **119,** 69–88 (1953).
27. Solomon, J. A. & Pelli, D. G. The visual filter mediating letter identification. *Nature* **369,** 395–397 (1994).
28. Majaj, N. J., Liang, Y. X., Martelli, M., Berger, T. D. & Pelli, D. G. The channel for reading. *J. Vision* [online] **3** ⟨http://www.journalofvision.org/⟩ (2003).
29. Selfridge, O. Pandemonium: a paradigm for learning. *Symposium on the Mechanization of Thought Processes* 513–526 (HM Stationery Office, London, 1959).
30. Johnston, J. C. A test of the sophisticated guessing theory of word perception. *Cogn. Psychol.* **10,** 123–153 (1978).

···

Control of dynamic CFTR selectivity by glutamate and ATP in epithelial cells

M. M. Reddy* & **P. M. Quinton***†

* *Department of Pediatrics, UCSD School of Medicine, University of California, San Diego, La Jolla, California 92093-0831, USA*
† *Division of Biomedical Sciences, University of California, Riverside, California 92521, USA*

Cystic fibrosis is caused by mutations in cystic fibrosis transmembrane conductance regulator (CFTR), an anion channel[1]. Phosphorylation and ATP hydrolysis are generally believed to be indispensable for activating CFTR[2]. Here we report phosphorylation- and ATP-independent activation of CFTR by cytoplasmic glutamate that exclusively elicits Cl^-, but not HCO_3^-, conductance in the human sweat duct. We also report that the anion selectivity of glutamate-activated CFTR is not intrinsically fixed, but can undergo a dynamic shift to conduct HCO_3^- by a process involving ATP hydrolysis. Duct cells from patients with ΔF508 mutant CFTR showed no glutamate/ATP activated Cl^- or HCO_3^- conductance. In contrast, duct cells from heterozygous patients with R117H/ΔF508 mutant CFTR also lost most of the Cl^- conductance, yet retained significant HCO_3^- conductance. Hence, not only does glutamate control neuronal ion channels, as is well known, but it can also regulate anion conductance and selectivity of CFTR in native epithelial cells. The loss of this uniquely regulated HCO_3^- conductance is most probably responsible for the more severe forms of cystic fibrosis pathology.

The molecular structure of CFTR is much more complicated than most ion channels. CFTR combines a regulatory domain consisting of numerous phosphorylation sites with two nucleotide-binding domains capable of ATP hydrolysis[3] flanked by six transmembrane domains on each side[4]. Consistent with this structure, a consensus has evolved that the phosphorylation by protein kinase A and hydrolysis of ATP are essential for activating CFTR Cl^- channel[2]. However, several discordant observations raised questions as to whether these requirements are absolute. We previously reported that CFTR Cl^- channels in the human sweat ducts are constitutively