BRILL

Seeing and Perceiving 23 (2010) 349–372

brill.nl/sp

Magnitude of Perceived Change in Natural Images May Be Linearly Proportional to Differences in Neuronal Firing Rates *

David J. Tolhurst [1,**], **Michelle P. S. To** [1], **Mazviita Chirimuuta** [1], **Tom Troscianko** [2], **Pei-Ying Chua** [1] **and P. George Lovell** [2]

[1] Department of Physiology, Development & Neuroscience, University of Cambridge, Downing Street, Cambridge, CB2 3EG, UK

[2] Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, UK

Received 11 September 2009; accepted 22 July 2010

Abstract

We are studying how people perceive naturalistic suprathreshold changes in the colour, size, shape or location of items in images of natural scenes, using magnitude estimation ratings to characterise the sizes of the perceived changes in coloured photographs. We have implemented a computational model that tries to explain observers' ratings of these naturalistic differences between image pairs. We model the action-potential firing rates of millions of neurons, having linear and non-linear summation behaviour closely modelled on real V1 neurons. The numerical parameters of the model's sigmoidal transducer function are set by optimising the same model to experiments on contrast discrimination (contrast 'dippers') on monochrome photographs of natural scenes. The model, optimised on a stimulus-intensity domain in an experiment reminiscent of the Weber–Fechner relation, then produces tolerable predictions of the ratings for most kinds of naturalistic image change. Importantly, rating rises roughly linearly with the model's numerical output, which represents differences in neuronal firing rate in response to the two images under comparison; this implies that rating is proportional to the neuronal response.

Keywords

Natural images, ratings, V1 model, transducer function

* This article is part of the 2010 Fechner collection, guest-edited by J. Solomon.

** To whom correspondence should be addressed. E-mail: djt12@cam.ac.uk

DOI:10.1163/187847510X532676

1. Introduction

Fechner and Weber suggested that, in many sensory domains, there is a geometric scale of sensory intensity perception and that this might arise from a compressive relation between neuronal response magnitude and stimulus intensity (Murray and Ross, 1988; Ross, 1995). These pioneering psychophysical ideas may not have led to a perfect unifying Law of Sensation (Stevens, 1961) but the proposals about psychophysical magnitude discriminations have underlain many of the important and *testable* comparisons between psychophysical performance and single neuron sensory physiology. Fechner and Weber might have supposed that 'delta-C is directly proportional to C', implying a logarithmic transducer function between neuronal response and contrast, but careful psychophysical measurements have suggested other compressive transducer shapes, and explanations have been sought in the actual behaviour of sensory neurons. This has been especially the case for the study of sinewave grating contrast discrimination (Boynton *et al.*, 1999; Chirimuuta and Tolhurst, 2005a; Foley, 1994; Goris *et al.*, 2009; Legge and Foley, 1980; Watson and Solomon, 1997). Measurements of how visual neurons respond to contrast (Albrecht and Hamilton, 1982; Heeger *et al.*, 2000; Sclar *et al.*, 1990; Tolhurst *et al.*, 1981, 1983) can inspire quantitative models of how a whole organism discriminates contrast stimuli (Watson and Solomon, 1997). In this paper, we extend such ideas to consider how the responses of visual cortex neurons to simple stimuli can help us understand how people perceive changes in natural images.

Since the earliest electrophysiological recordings of the responses of sensory neurons and the discovery that intensity is coded as action potential frequency (Adrian and Zotterman, 1926), it has been an important question how neuronal response properties relate to human psychophysical performance (e.g., Barlow and Levick, 1969; Borg *et al.*, 1966; Parker and Newsome, 1998; Werner and Mountcastle, 1965). The way that neuronal response depends upon stimulus intensity has been of particular interest, since it bears direct comparison with the well formulated psychophysical ideas of Fechner and Weber (e.g., Chirimuuta and Tolhurst, 2005a, b; Tolhurst *et al.*, 1983, 1989; Werner and Mountcastle, 1965). As early as 1931, B. H. C. Matthews studied the muscle spindle (a receptor which increases its activity when the muscle is stretched) and he reported: 'If the frequency [of action potential firing]... be plotted against the logarithm of the load, the points lie very nearly on a straight line.... It has long been held that, as a stimulus increases in geometric progression, the sensation increases in arithmetic progression (Fechner's Law)'. That is, he suggested that the logarithmic relation between response and load (i.e., the tension in the muscle's tendon) might underlie Fechner's Law of Sensation. While this looks like a pleasingly straightforward relation between neuronal behaviour and psychophysical performance, the observation points to a number of complications in trying to compare overall psychophysical performance with the behaviour of single neurons.

1.1. What Is the Appropriate Measure and Range of Stimulus Intensity?

The later study of the muscle spindle (P. B. C. Matthews, 1972) illustrates important caveats in the search for Laws of sensory coding or perception. First, it turns out that the response of the muscle spindle is determined by muscle *length* and not by *tension*, and it also turns out that tension is logarithmically related to length. Thus, response is actually *linearly* proportional to length, the appropriate measure of stimulus intensity in this case. In trying to deduce relations between neuronal response or psychophysical magnitude estimation and stimulus intensity (Stevens, 1961), it is clearly important to understand what system of measurement is appropriate for the stimulus intensity. Indeed, it may sometimes be the case (as in our present study) that there is no obvious single-dimensional metric to use.

The behaviour of the muscle spindle illustrates a second caveat: while it is a valid experimental strategy to study a system under the widest range of conditions, it should be noted when, in everyday life, that system is subject to only limited parts of the potential range. It is important to recognise the *natural intensity range* and to ask whether the system's behaviour is the same within the natural range as it is overall. Muscle spindle response is linearly proportional to muscle length (as said above), but only for very small stretches. While this seems, at first, to lessen any interest in a linear response range, it is compatible with the observation that, *in situ*, the length of a muscle is actually able to change rather little (about 5–10%), even though the joints be fully flexed or fully extended.

In this paper, we study human perception of naturalistic changes in digitised photographs of everyday scenes: i.e., natural images. In vision, we may be exposed at different times to stimuli that vary by many log units in intensity (mostly at different times of day). Stevens (1961) found that the power law relation between psychophysical magnitude and stimulus intensity for bright spots had an exponent of about 0.33 over a large range of intensities. However, when viewing natural scenes, the range of intensities in any one scene is much more modest. More generally, it is a tenet of vision science that the appropriate measure of stimulus intensity should be *contrast*, a measure of the intensity of an object *relative to the average level* (Enroth-Cugell and Robson, 1966; Troy and Enroth-Cugell, 1993) and, in a typical natural scene, the contrast is mostly low and ranges over only 2 log units (Brady and Field, 2000; Clatworthy *et al.*, 2003; Lauritzen and Tolhurst, 2005). The relation between perceived magnitude and contrast has a different power (0.65–1.0) from that reported by Stevens for intensity (Biondini and De Martelli, 1985; Cannon, 1979; Gottesman *et al.*, 1981; Peli *et al.*, 1991). It has also been questioned whether the appropriate metric might actually be contrast *energy*, the contrast squared (Solomon, 2009). As with the example of the muscle spindle, this visual example shows the importance of considering the appropriate intensity metric and a natural range of stimuli.

Simply equating neuronal response magnitude with the substrate for human magnitude estimates raises a third caveat. It is well accepted that human detection thresholds or intensity discrimination limens involve statistical judgments about

changes in an 'internal variable' (Green and Swets, 1966). It is a common observation in sensory neurophysiology that neuronal response variability (or 'noise') increases with response magnitude and presumably with stimulus intensity (Matthews and Stein, 1969; Tolhurst *et al.*, 1981, 1983, 2009; Werner and Mountcastle, 1965). Any model of how psychophysical judgments depend upon neuronal behaviour should include knowledge not only of how neuronal response magnitudes depend upon intensity, but also of how the 'response noise' changes (we shall consider this point in the Discussion).

1.2. Is a Single-Neuron Model Appropriate?

While it is attractive that overall psychophysical performance seems to relate to the response-intensity relations of single sensory neurons, it is obviously the case that neurons will not be active in isolation: overall performance must be the result of activity in many (probably disparate) neurons. Take, for instance, Weber's classical experiment on the ability to discriminate the weights of objects (Brodie and Ross, 1984). Suppose that we hold our upper arms by our sides, our forearms horizontally with our palms upright and that we hold two different weights on our two palms. It may be that there are muscle spindles (compare B. H. C. Matthews, 1931) in the biceps whose stretch and responses will depend on the weights on the palms. However, there are very many different kinds of sensory receptor in the arm (including the elbow or wrist), which must surely be involved (spindles in fore-arm and upper-arm muscles, receptors in joints, and pressure receptors in the skin of the hand). The magnitude of response given by a muscle spindle in response to a given stretch also depends upon efferent outflow from the nervous system and is affected by the amount of force needed to keep the forearms horizontal (P. B. C. Matthews, 1972). Finally, it is likely that judgements about the relative weights in the two hands will involve some deliberate 'testing' motor activity, determining how much extra force is needed to lift the objects a small distance (Brodie and Ross, 1984). In all, the neural activity is complex and disparate even for a 'simple' judgement of which of two stimuli is the more intense. How much more complex when the stimuli vary in multiple dimensions that are not easy to quantify individually?

A true picture of how perceptual magnitude relates to neural responses requires a comprehensive description of the individual contributions of all the disparate sensory signals involved, and hypotheses about how those many signals are weighted and combined to give a single judgement. In fact, just such an enterprise has been ongoing for 25 years in vision science (Daly, 1993; Foley, 1994; Lovell *et al.*, 2006; Lubin, 1995; Parraga *et al.*, 2005; To *et al.*, 2009, 2010; Watson, 1987; Watson and Ahumada, 2005; Watson and Solomon, 1997). Using the immense amount of quantitative data on single V1 neuron responses and human channel behaviour, it has been possible to build computational models of how millions of neurons respond to visual stimuli, and then to compute how the population behaviour differs in response to different stimuli. Such models are used to understand the neural contributions to 'simple' decisions about stimulus contrast (an intensity metric in the

line of Weber and Fechner), and they have also been used as the basis of image quality metrics to describe the degree of perceived corruption in images that have been subject, say, to JPEG compression (e.g., Daly, 1993; Lubin, 1995).

We have been investigating human suprathreshold perception of changes in natural images (To *et al.*, 2008, 2009, 2010): e.g., changes in the shapes, colours or numbers of objects in a scene. The great variety of image changes in our experiments (e.g., changes in hue, blur, object size or posture, number of objects in view) means that there is no obvious single physical stimulus metric against which we can compare the observers' ratings to all those change types, separately or in combination; a model of the responses of millions of V1 neurons is an attempt to unify the data. In this paper, we shall examine whether such a model helps us to understand the relationship between psychophysical judgements and neural population response for suprathreshold changes in natural visual images. We have already applied a V1-based model to the results of a ratings experiment with a disparate set of images constructed with an unsystematically chosen variety of image changes (To *et al.*, 2010); we argued that some of those image changes (e.g., changes in facial expression) might not be amenable solely to V1 modelling. Here, we extend the V1 modelling to show explicitly how a single model with few parameters can link 'classical' psychophysical experiments on contrast discrimination ('dipper' functions with sinewave gratings) to our experiments which measure the perceived magnitudes of changes in, for example, the shapes, colours, locations and numbers of objects in natural scenes. We shall discuss how the 'transducer function' (Legge and Foley, 1980), which we model as underlying the dipper function, is related to the population response of V1 neurons. We ask particularly whether the model gives a straight-line fit to our ratings data on natural image differences. We concentrate on a newer set of experimental results, obtained with stimuli in which the various changes (e.g., in object size, location and colour) are systematically changed (separately and in combination) to provide a well-distributed range of rating data to challenge the model.

2. Methods

2.1. Coloured Natural Image Stimuli and Magnitude Estimate Ratings

Our methods for constructing and presenting stimuli are given in detail by To *et al.* (2008, 2010).

In a first experiment (To *et al.*, 2008), 294 image pairs were made from a small number of parent images. Six parent images led to 48 variants each. An image could differ from the parent along one dimension, along a second dimension or along both together. Including 'no change', there were 7 steps along each dimension (the steps were intended to be of equal perceptual magnitude), giving 7×7 variants in total. Variants could differ in the locations of objects within the image (e.g., Fig. 1A, B), the sizes or colours of objects, or in the intensity of shadows (Fig. 1B). The 294

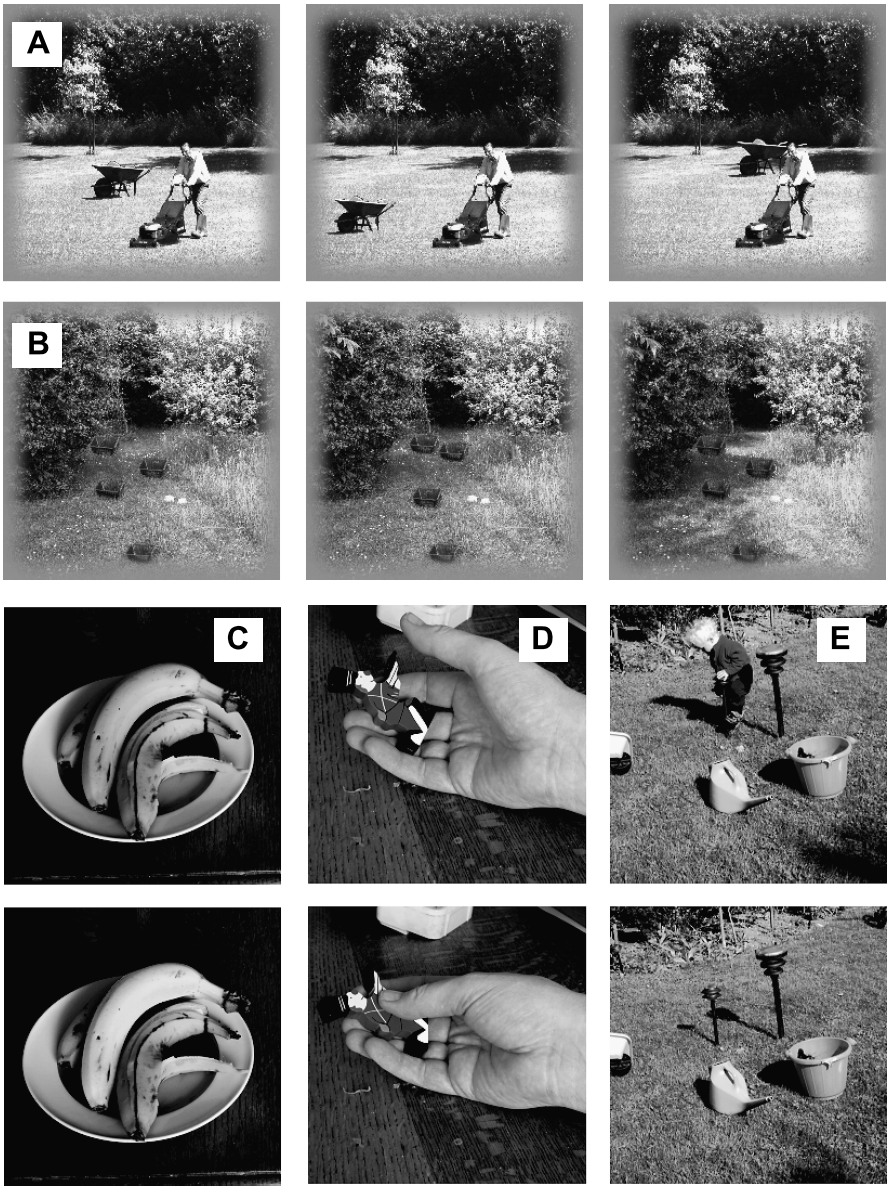


Figure 1. Monochrome representations of some of the kinds of image pair used in our experiments. (A) and (B) from the 'garden scene' series. The left-hand images show two of the parent images in the experiments, while the middle and right-hand images show variant images against which the parents could be compared. There were 48 variants of each. (A) Shows two variants that differ in the magnitude of a single change type, while (B) shows variants that differ in 2 different ways. (C–E) From the 'varied pairs' series; the upper stimulus is the parent, and the lower image is one of 5 variants in the experiments. (C) Shows a colour change; (D) shows a shape change; (E) shows an item disappearing. The 'colour change' was achieved by changing the hue and the saturation of one banana, using code written in Matlab (The Mathworks). 'Shape' and 'appearance' changes used time-lapse photography. For details and coloured examples, see To *et al.* (2008, 2010).

images could be presented upright or inverted, in random order, giving 588 pairs altogether. We have not previously applied a V1-based model to these data.

In a second experiment (To *et al.*, 2010), 900 pairs of images were made from a wide variety of coloured photographs of natural scenes, covering subject matter such as animals, plants, people, man-made objects, landscapes, still-lifes and garden scenes. Some image pairs could be made by taking one natural photograph and using some kind of image processing technique to change the colour (hue and/or saturation) of all or part of the scene (e.g., Fig. 1C; coloured examples are given in To *et al.*, 2008). Images could also be blurred or sharpened. Many image pairs were made from a pair of photographs of the same scene. In the time between photographs, the shape or arrangement of objects in the scene may have changed (e.g., Fig. 1D), or an object may have appeared or disappeared (e.g., Fig. 1E), or the natural lighting and shadowing may have changed due to change in the time of day or the weather. Some image pairs were made by combining the natural shape changes with image-processed colour or blur changes. There were 180 parent images, each paired against 5 variants.

The images were 256 by 256 pixels, 3.2 deg square surrounded by uniform grey in a larger display. The stimuli were presented through a ViSaGe system (Cambridge Research Systems) so that we had precise control and knowledge of the luminance of each pixel. In a given trial, the observer had to compare two related images; one was a parent image and the other was a variant of it. The observer viewed a small spot in the centre of the display, and then the images in a pair were presented sequentially. The first image was presented for 833 ms followed by a 83 ms interval when the screen was uniform grey apart from the fixation spot. The second image was then presented for 833 ms, followed by a 83 ms grey interval and a 833 ms re-presentation of the first image. The fixation spot was extinguished while the images were present, but the observers were instructed to view the centre of the images and not move their gaze about. The observer then gave a numerical rating of the perceived difference between the images. Every 10 trials, one particular image pair was presented (a picture of a red flower where the difference was in colour saturation); the numerical difference between this standard pair was set to '20', and observers were instructed and trained to use a ratio scale to rate any kind of difference in any other image pair with respect to this standard. The observers were sometimes surprised by this instruction at first, but after a little practice, they mostly reported being very comfortable with the idea that they *could* rate any perceived change against, say, a colour saturation change. However, this does raise the question whether the observers really did maintain a single rating standard or whether they might have unintentionally held slightly different standards for different image change types. Observers could give a rating of '0' if they perceived the images in a pair to be identical; they had been told that some of the image pairs might be identical, but they were not told what proportion. There was no upper limit set for the ratings. At the end of the experiments, each of the observers' ratings was

divided by the median value for that observer, and the normalized ratings for each pair were averaged across observers.

This stimulus presentation protocol is similar to those that might produce 'change blindness' (Simons and Rensink, 2005). However, apart from image pairs that differed in the detailed organizations of textures or in small movements of objects (see below), the observers did not seem to be subject to change blindness. The observers had had sufficient training and practice (see To *et al.*, 2010) with image pairs like those to be presented in the main experiment that, we presume, they were clearly aware of what sorts of image change to expect.

2.2. Contrast Discrimination Dippers

Our methods for stimulus presentation and the staircase procedure for obtaining contrast thresholds have been described in detail before (Chirimuuta and Tolhurst, 2005a). *Pedestal* images were presented at a variety of contrasts, defined as dB attenuation from the maximum (when the brightest pixel in the image was double the surrounding grey of the display and the darkest pixel had a nominal value of zero). The pedestal image was a 6 deg square image derived either from a monochrome photograph of a street scene (Fig. 2A) or a $1/f$ random noise pattern (Fig. 2C) or versions of these that had been notch filtered or bandpass filtered with 1 octave wide filters. The *increment* was added to the pedestal and its contrast was adjusted in a 2AFC staircase to determine the increment threshold — the increment and pedestal were presented on alternate frames (frame rate 120 Hz) and their contrasts were controlled separately using pseudo-15 bit LUTs. The increment was a Gaussian-weighted patch of the street scene, the $1/f$ pattern or a filtered variant; the spread of the Gaussian was 0.38 deg. Eight different combinations of increment and pedestal were studied, with 11 different pedestal contrasts in each. In an experimental session, the staircases for 5–6 pedestal contrasts would be randomly interleaved.

In a trial, the pedestal would be presented alone in one 100 ms interval and the pedestal plus the increment in the other interval. The increment was assigned to the first or second interval at random on each trial. In response to the observer's choices, the increment contrast was increased or decreased, and the staircase generally stabilized at a contrast near to the point where the observer would correctly choose the increment interval on 75% of trials. Threshold was calculated by fitting an error function to the psychometric function that resulted from 100–200 trials.

2.3. A V1 Based Model of Visual Discrimination

Our model is based on that of Watson and Solomon (1997) which is used to explain detection thresholds for monochrome grating stimuli. We have tried to extend that model to encompass coloured stimuli (see Lovell *et al.*, 2006; To *et al.*, 2010) and suprathreshold decisions. We have implemented a variety of models with different receptive field shapes and interactions (To *et al.*, 2010) and here we describe the one that yielded the best match to the psychophysical dipper data: a phase-invariant 'complex cell' model.

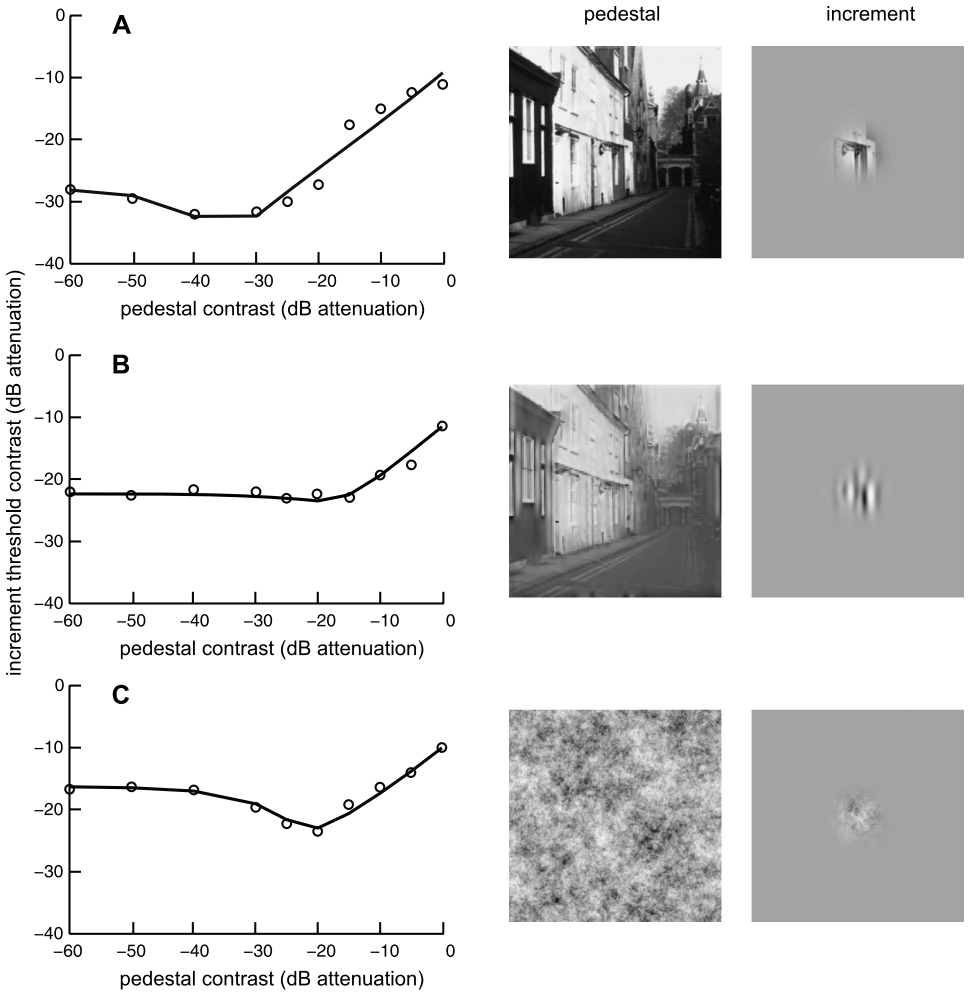


Figure 2. Contrast 'dipper' experiments that were used to optimize the numerical parametersof the V1 model. In all, 8 experiments were examined, and a single set of numerical parameters fitted to them, and 3 of those 8 are summarized here. (A) The circles show the average discrimination thresholds of 2 observers in an experiment where a small patch of the image was added to the centre of a monochrome photograph of a street scene. The line is the dipper calculated from the V1 model with the optimized parameters. (B) The same except that the full street scene has been notch filtered to remove vertical components near 16 cycles per picture, while the test patch was band-pass filtered around vertical 16 c/picture. (C) The same except that the full 'picture' is a $1/f$ filtered random noise pattern; the small central patch is of the same filtered noise.

Briefly, the model consists of 'simple-cell' receptive fields with Gabor profile — 6 orientations by 5 spatial frequencies by two spatial phase symmetries. The fields are elongated and are not self-similar: the spatial frequency and orientation tuning gets sharper as spatial frequency is increased (Tolhurst and Thompson, 1981; Yu *et al.*, 2010). The bandwidths were graded with optimal frequency: for fields with op-

tima of 1.25, 2.5, 5, 10, 20 cycles per degree, the frequency bandwidths were 2.12, 1.43, 0.93, 0.64, 0.45 octaves (full width at half height) and the orientation bandwidths were 43.4, 34.5, 22.8, 17.7, 11.6 degrees (full width at half height). These values lie in the ranges found for single neurons and deduced for psychophysical channels. The sensitivities of these frequency bands within the model were weighted according to typical observers' contrast sensitivity for gratings of those frequencies. There are, in fact, three of each field type to deal with image colour: a 'luminance' detecting field, and an isoluminant red–green opponent and a blue–yellow opponent field based on the MacLeod–Boynton transformation (Lovell *et al.*, 2006).

Each field must be represented at every spatial location in the images and so the field templates must be convolved with the images. The first image in the pair is first convolved with all the different field types, giving a set of values proportional to image luminance. These linear responses are divided by the local mean luminance values to give *contrast* responses. We follow the method of Peli (1990): for each spatial frequency band, the image is also convolved with a 2D Gaussian blob having the same spread as the Gaussian envelope of the Gabor fields. The 'linear response' of each Gabor field is divided by the 'linear response' of the matched Gaussian blob whose field is centred on the same point in space.

The r.m.s. is taken of the responses of paired odd-symmetric and even-symmetric 'simple cells' to give phase-invariant 'complex cell' responses. The contrast output of each 'complex cell' is weighted according to estimates of a human observer's contrast sensitivity for luminance or isoluminant sinusoidal gratings of the field's centre spatial frequency, orientation and eccentricity from the fovea. We measured foveal contrast thresholds for luminance gratings of appropriate orientations and spatial frequencies, and deduced the sensitivities within the model so that the model would explain those thresholds; we estimated thresholds for chromatic gratings from Mullen (1985) and Mullen and Kingdom (2002). The spatial fall-off in sensitivity from the centre of the fovea was modelled from Pointer and Hess (1989) and Mullen and Kingdom (2002). For luminance gratings the sensitivity falls by a factor of 10 in about 40 cycles of the grating along the vertical meridian and in about 60 cycles along the horizontal meridian (Pointer and Hess, 1989; Robson and Graham, 1981). The sensitivity for isoluminant gratings falls off much more quickly with eccentricity (Mullen and Kingdom, 2002): a factor of 10 in only 8 cycles for low spatial frequency RG gratings and in 15 cycles at higher frequencies. The sensitivity for BY isoluminant gratings falls off by a factor of 10 in about 25 cycles. The point of maximum sensitivity was, of course, in the centre of the target image.

In the simplest model, the second image would be similarly processed and the responses to the image pair would be compared neuron by neuron. However, realistic models of V1 incorporate threshold behaviour, and non-linear suppressive interactions between neurons (e.g., Blakemore and Tobin, 1972; Heeger, 1992). The simple

contrast responses of the neurons must be transformed (following Legge and Foley, 1980) by a sigmoidal transducer function of the form (see Fig. 3):

$$Response = \frac{|Contrast|^p}{1 + W_N \cdot Normalise^q + W_S \cdot Surround^r}. \quad (1)$$

The numerator with power p gives a positive acceleration or threshold; the terms in the divisor represent two different kinds of inhibitory interaction between neurons and their effect is to cause the response to become compressed at higher contrasts. Divisive normalizing nonlinearities have been established neurophysiologically and psychophysically (Bonds, 1989; Foley, 1994; Heeger, 1992; Watson and Solomon, 1997). We have modeled two forms of suppressive signal, treating them as distinct since neurophysiology has usually shown them to have different properties. However, Cavanaugh *et al.* (2002) do suggest that the two suppressive phenomena may grade into each other. First, we model a *spatially localized*, *non-specific suppression* (Heeger, 1992); we implemented the normalizing signal for all the neurons at a given point in the image as the sum of the responses of all the neurons with fields centred at that point; the absolute value of each suppressing neuron's response was raised to a power q before the sum. Thus the spatially-localized signal is summed across all spatial frequencies and orientations, and the same divisive signal is applied to all neurons at that point. We have also modeled *surround suppression that is stimulus specific* (Blakemore and Tobin, 1972; Cavanaugh *et al.*, 2002; Maffei and Fiorentini, 1976); this is necessary to explain the forms of contrast discrimination functions for gratings of different geometry (Meese, 2004). Here, the suppressive signal applied to a neuron at a given location in the image is derived only from neurons of the same optimal spatial frequency and orientation, but whose fields are centred in a blurred annulus around the neuron being suppressed. The absolute values of the responses of these suppressing cells are raised to a power r, then summed and weighted according to a blurred annulus around the inhibited neuron:

$$Surround\_strength_f = d \cdot \mathrm{e}^{-d^2/(2 \cdot rad_f^2)}, \quad (2)$$

where d is distance from the center of the suppressed field, and rad_f is the radius of the annulus, which is proportional to the spatial period of the carrier sinewaves of the Gabor fields. Thus, we have two suppressive signals: one is spatially very localized but is diffuse in spatial frequency and orientation (Heeger, 1992), whilst the other is specific to spatial frequency and orientation but diffuse in space (Blakemore and Tobin, 1972). Watson and Solomon (1997) used one suppressing signal which is potentially more diffuse in space than ours and is partly specific for orientation or frequency; Cavanaugh *et al.* (2002) would suggest that the degree of stimulus specificity should vary with distance from the centre of the suppressed neuron's receptive field.

The transformed responses of the millions of neurons to one image are subtracted from the transformed responses to the other image in the pair, and the millions of differences are pooled by Minkowski summation of the absolute values (Watson

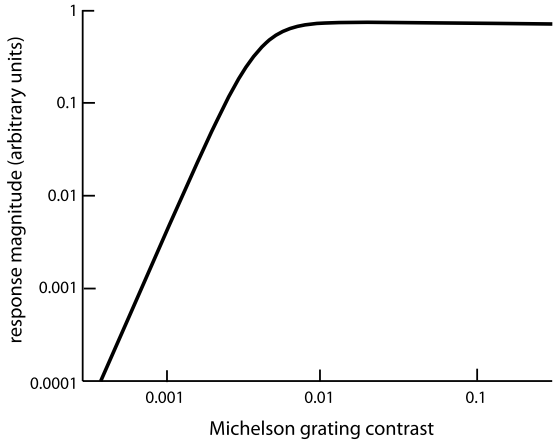
and Solomon, 1997) with exponent m. The luminance, red–green and blue–yellow planes were processed entirely separately, and the cues they provided to image difference were finally combined only at this Minkowski summation stage. This gives a *single number* which is the model's prediction of the observer's magnitude estimate. Although the images may have differed in many, disparate ways along different dimensions, the whole dataset of 588 or 900 naturalistic images pairs is summarized by giving a single numerical output value to each image pair.

The 7 model parameters (5 in equation (1), rad_f and m) were adjusted to give the best overall fit to the 88 data points in 8 contrast discrimination experiments with monochrome images (Fig. 2). The iterative search minimized the sum of squares deviation between model prediction and experimental thresholds. However, the fit with just those 7 parameters was not satisfactory: some model dippers were slightly displaced above the experimental data and some slightly below. We added an extra private parameter to the fits for 6 of the dippers to allow the model to slide each model dipper to better fit the data. The 6 extra parameters increased or decreased the grating contrast thresholds underlying the model by up to $+/-2$ dB, independently for each dipper. These shifts may represent day-to-day differences in observer sensitivity or differences in the adaptational state caused by slight differences in the contrast energy in the stimuli making up the different dipper experiments.

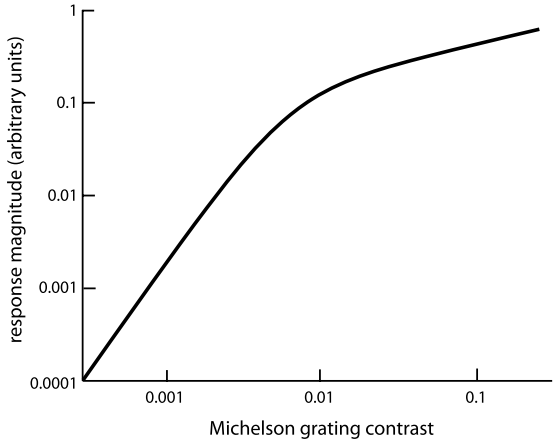
3. Results

The 7 numerical parameters of the V1 model (plus the 6 threshold adjustment parameters) were sought by iteratively searching to minimize the sum of squares deviation between the results of 8 contrast discrimination 'dipper' experiments (11 thresholds each) and the predictions of the model for the 88 threshold measurements that comprised the experiments. The mean sum of squares error (MSE) between model and experimental data point was 2.21 dB squared, compatible with estimated standard errors on the threshold measurements of 1–1.5 dB. Figure 2 shows the experimental results and the model fit for 3 of those 8 dippers; these fits are representative of the MSE across the whole data set. The images beside the graphs show

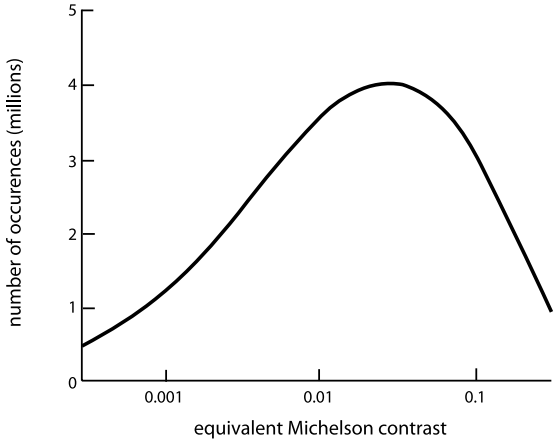
Figure 3. (A) The sigmoidal transducer function resulting from our V1-based model of contrast discrimination in monochrome naturalistic stimuli (Fig. 2). The transducer shows the response of a 'neuron' to a sine-wave grating; the neuron's receptive field matches the orientation and spatial frequency of the grating, and is in the centre of the grating (the 'fovea'). (B) The transducer function reported by Legge and Foley (1980) to describe their 'dipper' functions with sine-wave gratings. The graph of (A) may seem to have a threshold at low contrasts, while that of (B) has a small positive acceleration. The different appearance only reflects how the different graphs intersect the identical log–log axes in parts (A) and (B). (C) A histogram of the equivalent luminance contrasts in a selection of our naturalistic stimuli. Equivalent contrast is calculated from the first stage of the V1 model: the image is convolved with a receptive-field template and the resultant convolution is divided by the local mean luminance; that value is then calibrated against the model's initial response to sine-wave gratings of known contrast and the appropriate spatial frequency (Tadmor and Tolhurst, 2000).


response magnitude (arbitrary units)
1
0.1
0.01
0.001
0.0001
0.001
0.01
0.1
Michelson grating contrast

(A)


response magnitude (arbitrary units)
1
0.1
0.01
0.001
0.0001
0.001
0.01
0.1
Michelson grating contrast

(B)


number of occurences (millions)
5
4
3
2
1
0
0.001
0.01
0.1
equivalent Michelson contrast

(C)

the pedestal stimulus, based either on a photograph of a street scene or an image of $1/f$ filtered noise. The 8 dippers used these pedestals or bandpass or notch-filtered (e.g., Fig. 2B) versions of these. The increments were small Gaussianly-weighted central patches of the main images or filtered versions of them. The best-fitting exponents p, q and r in equation (1) were 4.23, 3.45 and 3.86 with a Minkowski summing exponent m of 2.16. These values are comparable with those in other studies, primarily of the visibility of sinewave gratings (Foley, 1994; Foley *et al.*, 2007; Watson and Ahumada, 2005; Watson and Solomon, 1997). The weights W_N and W_S were 0.0504 and 0.779, with a surround radius rad_f of 2.14 periods.

For the 'dipper' in Fig. 2A, the y-axis intersection (-28 dB) and the pedestal contrast at the depth of the dip (about -35 dB) are lower than for the other two illustrated 'dippers'. This is likely because the street scene (with its regularly spaced vertical door and window frames) contains a band of higher contrast energy than do the other stimuli. The biggest equivalent Michelson contrast (see below, Fig. 3C) in the street scene is 0.56, but is only 0.19 in the $1/f$ filtered noise (an approximately 10 dB difference).

Six of the 7 numerical parameters determine the form of the Naka–Rushton transducer function (the 7th is the Minkowski summing exponent m). Figure 3A plots the transducer function for the optimized set of parameters; it shows the result of equation (1) for the single 'neuron' giving the largest response when the stimulus is a sinusoidal grating covering the full spatial extent of the modeled x, y space. Figure 3B shows the transducer given by Legge and Foley (1980) for their seminal description of contrast discrimination of sinewave gratings. They used a simple Naka–Rushton formulation with p of 2.4 and q of 2.0. The transducers in Fig. 3A and 3B are not identical, but the inflexion is at a remarkably similar contrast in the two.

For interest, Fig. 3C tries to indicate approximately where in the non-linear transducer range our model is operating for natural image stimuli. The histogram shows the frequency with which the many 'neurons' in the model 'saw' the stimulus as containing a feature giving the same magnitude of response as their favoured sinewave grating — the equivalent Michelson contrast, which is the contrast of the optimal sinewave grating that would evoke the same response as did that location in the natural image (Lauritzen and Tolhurst 2005; Tadmor and Tolhurst, 2000). These responses are calculated at an early stage of the model: after the convolution with receptive field templates and division by the mean luminance, but before application of the suppressive nonlinearities. The modal value of 'contrast in natural images' is low, as we have shown before, and is close to the inflexion in the non-linear transducer functions.

The model, with exactly the same parameters as shown for the contrast-discrimination dippers of Fig. 2, was applied to the coloured image pairs in two experiments where observers gave magnitude estimation ratings of the perceived difference between images in each pair. The 7 model parameters were, of course, derived for monochrome images (Fig. 2), and we used the same parameter values

for the MacLeod–Boynton RG and BY planes of the images, which we have used to investigate perception of chromatic changes (see To *et al.*, 2010, for detailed discussion). The model treated the Luminance plane and the RG and BY planes entirely separately; the only differences in processing were in the luminance or isoluminant grating sensitivities fed into the model and the ways in which these fell off with eccentricity (Mullen and Kingdom, 2002).

Figure 4A plots the observers' ratings for the 294 'garden scene' image pairs against the output of the model (which is in arbitrary units). Each image pair was presented once upright and once upside-down; we were interested whether high-level cognitive cues might have contributed to the ratings and we supposed that image inversion might frustrate such cues. The results for upright and inverted are very similar. The solid line shows the robust linear regression through the graph (r is 0.83; n of 588). It is worth noting that the ratings were correlated with the pixel-by-pixel r.m.s. difference between the images with r of 0.60. The dashed line shows the regression with an added quadratic term. Although, the modest curvature of the quadratic regression does not seem to add much to the overall relation, addition of the quadratic term did have a highly significant effect on the residuals (F of 75.7).

Figure 4B plots the ratings for the 900 'varied image pairs' against the output of the model. There is clearly scatter in the fit of the ratings to the model ($r = 0.55$; $n = 900$) and, not surprisingly, there must be more to predicting an observers' ratings than understanding only the low-level coding processes of V1. While this correlation is low, it is substantially higher than the correlation between the ratings and the r.m.s. differences between the images in the pair ($r = 0.28$). However, the correlation is adequate enough that we can see that rating is roughly directly proportional to the model's output; adding a quadratic term (dashed line) gave insignificant improvement.

We have discussed in detail (To *et al.*, 2010) possible reasons why the V1-based model might give poor correlations for some kinds of image change. The higher correlation in Fig. 4A is probably because the 288 image pairs differed along only a few dimensions and because the stimuli were based on only 6 parent images so that each dimensional change was represented at 7 magnitudes. The 900 images pairs in the other set (Fig. 4B) were derived from 180 rather different parent images, and the pairs could differ in one or more disparate ways. We have argued (To *et al.*, 2010) that one particular kind of image change will be badly modeled: image changes where there are *small* changes in object location or where there are changes in *textures* of image parts (e.g., changes in the detailed arrangements of pebbles on a gravel path). The model, very literally, compares the two images with exact spatial precision, and this seems an unrealistic expectation of human vision. Figure 4C replots the data of Fig. 4B, after discarding the many image pairs where the major difference was in object textures or where objects moved only a small amount. The remaining ratings have a higher correlation with the model ($r = 0.65$; $n = 722$), but there is still work to explain all the variance in the observers' ratings. Again, adding a quadratic term to the regression did not give any significant improvement.

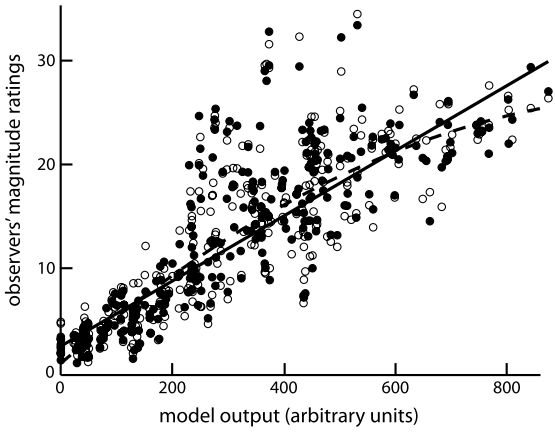
Overall, the 3 graphs in Fig. 4 seem to show that observers' ratings of the multidimensional and disparate changes in natural images tend to be directly proportional to the output of the model, based on the pooling of the supposed response magnitudes of cortical neurons. Quadratic regressions have only slight curvature or add nothing beyond a linear regression.

4. Discussion

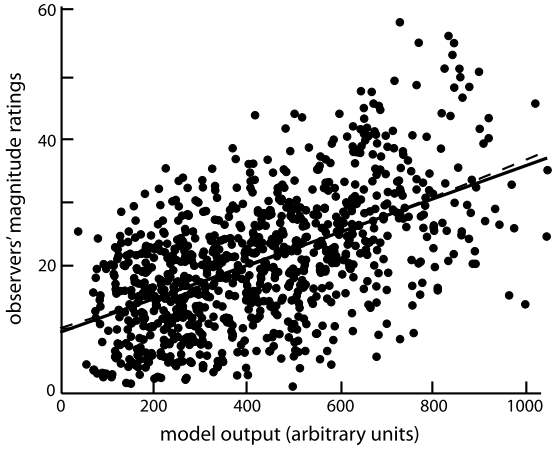
We have extended a quantitative model (Daly, 1993; Lubin, 1995; Watson, 1987; Watson and Solomon, 1997) of populations of V1 neurons to examine how it would fare with naturalistic stimuli (Rohaly *et al.*, 1997; To *et al.*, 2010). We have determined the numerical values of the model parameters by optimizing the model on a relatively straightforward experimental paradigm: contrast discrimination 'dippers' (compare Foley, 1994; Legge and Foley, 1980; Solomon, 2009; Watson and Solomon, 1997). Rather than being constructed conventionally with sinusoidal gratings, our 'dippers' were performed with a photograph of a natural scene, a $1/f$ random noise pattern (a surrogate for natural images whose power spectra are very roughly $1/f$) and various filtered versions of these. The model, therefore, was designed to give a good match to experiments of the same kind as considered by Weber and Fechner — measurements of the ability to distinguish stimuli of different intensity. We have then examined how the same model 'perceives' the suprathreshold differences between the coloured naturalistic images for which we have observers' magnitude estimate ratings. The images in these experiments covered a very large range of subject matter and there were many kinds of image change, which might be considered as the basic elements in many everyday visual tasks (see To *et al.*, 2010).

V1-based modeling (after Watson, 1987) has proved useful in developing metrics for assessing the quality of compressed or corrupted images (e.g., Daly, 1993; Lubin, 1995), and Rohaly *et al.* (1997) have also tried to model the visibility of targets in terrain scenes. We, too, are interested in applying vision research to explain the visibility of objects in natural scenes and the detectability of changes in those scenes. After five decades of fundamental psychophysical and neurophysiological research on the elements of visual processing and coding, we are in a position to ask whether all the careful and systematic experiments on, for example, sinusoidal gratings is enough to start to model everyday vision. While we cannot possibly say that

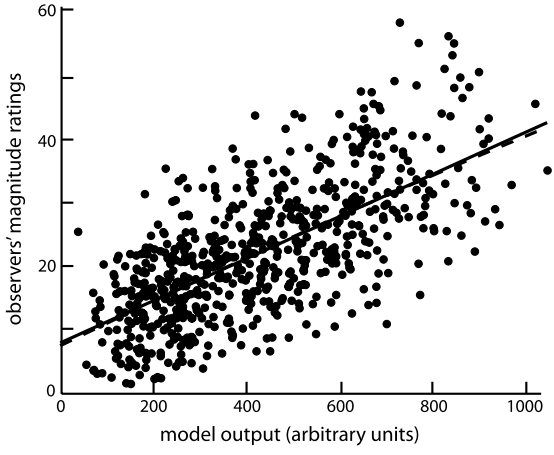
Figure 4. Plots of the experimental magnitude ratings against the output (in arbitrary units) of the V1 model optimized on contrast-discrimination dippers. The solid lines show robust least-squares regression lines, while the dashed lines have an added quadratic term. (A) The ratings for the 294 garden scenes are plotted against the model output ($r = 0.83$); each image pair was presented twice — once upright (filled symbols) and once upside-down (open symbols). (B) The ratings for all 900 'varied pairs' series are plotted against the model output ($r = 0.55$). (C) As for (B), except that a subset (722) of the 'varied pairs' is plotted ($r = 0.65$); stimulus images differing in small spatio-chromatic ('texture') changes have been discarded.


observers' magnitude ratings
0
10
20
30
0
200
400
600
800
model output (arbitrary units)

(A)


observers' magnitude ratings
0
20
40
60
0
200
400
600
800
1000
model output (arbitrary units)

(B)


observers' magnitude ratings
0
20
40
60
0
200
400
600
800
1000
model output (arbitrary units)

(C)

we have uniformly sampled the disparate multi-dimensional space of all possible natural images, we have considered a great variety of everyday scenes and everyday differences (see To *et al.*, 2008, 2010 for many examples). The correlations we have shown (Fig. 4) between human response and V1 model are promising, and suggest that it is not premature to study vision with natural images. We are also in a position to ask how different aspects of visual processing contribute to everyday vision and whether the detailed knowledge of V1 processing is sufficient. We have reported (To *et al.*, 2010) that, surprisingly, a model without the well-studied phenomena of divisive normalization fares little worse than a full model. We have been able to show that V1 modeling will have limits, since there are some scenes or scene changes where the models consistently fail to explain human performance, such as changes in facial expressions, shadowing or in textures (To *et al.*, 2010).

In one experiment, with a relatively limited range of images and image changes, the correlation between experiment and model output was 0.83; in the other experiment, with a greater range of subject matter in the images, the correlation was poorer. We have discussed elsewhere (To *et al.*, 2010) why the implementation of such models may fail to match an observer's perception of the magnitude of some kinds of image change. It should be noted that V1-inspired models are substantially better predictors of ratings than the pixel-by-pixel r.m.s. difference between images in the pairs. It should also be noted that we can develop 'better' models of the ratings by optimizing the model parameters on the rating experiment stimuli themselves, rather than on the 'dipper' data. The best correlations that we have obtained for the data in Fig. 4 are 0.86, 0.65 and 0.73 with a model in which the combination of numerical parameters is such that the transducer (compare Fig. 3A) is no longer monotonic. While such a re-optimized model may be useful in some other context, we wish to stress that the model we describe here directly links to the tradition of Weber and Fechner because it is optimized on intensity-discrimination data.

The pioneers of sensory science (Fechner, Weber, Adrian, and B. H. C. Matthews) supposed that a person's numerical magnitude judgment would be directly proportional to the response magnitude of appropriate neurons. We can argue whether sensory receptors or cortical neurons are the appropriate neurons to choose, and we can argue about the physical metrics used to define stimuli. We can surely accept that stimuli of greater magnitude must produce some greater internal neural response and that this leads to an increased magnitude of sensation and increased numerical rating values. However, it has been debated whether the internal magnitude of sensation need be directly proportional to neuronal response or whether numerical magnitude estimation ratings need be directly proportional to the internal magnitude of sensation (e.g., Gescheider, 1997). We instructed our observers to give ratings proportional to the magnitude of their sensations; in so far as the model gives an independent measure of image difference, the linear relation between ratings and model predictions implies that the observers did use an appropriate scale.

Thus, our present results suggest that observers' ratings do depend linearly on neuronal response levels. The rating of perceived complex differences in natural

scenes seems directly proportional to the numerical output of our V1 modelling which, we suppose, reflects differences in the magnitudes of neuronal responses to the two images under comparison. At first sight, it would seem that the perceived magnitude difference between two natural images is directly proportional to the difference in neuronal response to the two images, where response might be expressed as total number of action potentials generated during an image presentation. This presumes that the transducer function (Fig. 3A) underlying our modelling *is* a representation of how neuronal response magnitude increases with contrast. Boynton *et al*. (1999) have argued that the sigmoidal transducer function which is presumed to underlie contrast discrimination thresholds is the same shape as the relation between the V1 BOLD (fMRI) signal and contrast while Heeger *et al.* (2000), in turn, argue that the BOLD signal follows the relation between neuronal action potential rate and contrast.

However, the hypothesized sigmoidal transducer function (like equation (1); Fig. 3) does not simply describe the the relation between response amplitude and contrast for single V1 neurons. Individual V1 neurons each respond to very limited ranges of contrast, while the dynamic ranges of different neurons cover different contrast ranges (Albrecht and Hamilton, 1982; Sclar *et al.*, 1990; Tolhurst *et al*., 1981, 1983). The BOLD response is like the summed activity of many neurons whose threshold contrasts differ and whose responses saturate at different contrasts (Heeger *et al*., 2000). Thus, if we were to model the psychophysical transducer truly from the behaviour of V1 neurons, we would have to pool the responses of many neurons, and the transducer's shape would reflect the number of neurons responsive at each contrast as well as the shape of the response-contrast functions of single neurons (Chirimuuta and Tolhurst, 2005a; Clatworthy *et al*., 2003; Goris *et al*., 2009; Watson and Solomon, 1997).

Furthermore, psychophysical judgments are probabilistic and there is no explicit decision 'noise' within our model. We have modelled average performance rather than trial-by-trial variability, and so fixed noise is implicit as the fixed extra amount of 'response magnitude' in the transducer (Fig. 3A, B) that would be needed, on average, for discrimination. However, we need to recognize that the variability of neuronal responses as well as their magnitude is not fixed with contrast; our ability to discriminate contrasts will be affected by how response variability changes with contrast (Kontsevich *et al*., 2002). It is an often repeated observation that the variance of neuronal firing rates increases with increasing response level in visual cortex (Dean, 1981; Geisler and Albrecht, 1997; Snowden *et al*., 1992; Tolhurst *et al*., 1981, 1983; Vogels *et al*., 1989; Wiener *et al*., 2001). The standard errors of our ratings tend to be higher for higher rating values (To *et al*., 2010). Even if firing rate were linearly proportional to contrast, then we would still see something approaching Fechner's Law (despite the absence of a logarithmic transducer) because we would need bigger contrast increments at the higher contrasts to overcome the higher response variability (see Solomon, 2009). The transducer of Fig. 3 may not be simply a schematic of how pooled neuronal firing rate depends upon contrast;

it is an 'effective transducer' in that it fits the dipper data, implicitly incorporating any changes of variance (Chirimuuta and Tolhurst, 2005a; Goris *et al*., 2009; Kontsevich *et al*., 2002). It may, however, be that the dominant source of noise in decision processes lies elsewhere than in V1 and that it is less dependent on stimulus intensity.

In fact, the interpretation of the contrast dipper function is debatable (Georgeson and Meese, 2006; Kontsevich *et al*., 2002; Solomon, 2009). According to signal detection theory, appropriate contrast-dependent changes in response variance could as much give rise to the contrast dipper as could contrast-dependent changes in response magnitude; 'it could be argued that the transducer is not much of an explanation, simply an alternative way to describe the data' (Solomon, 2009). However, the effective transducer *does* look similar to the response *magnitude* relation implied by the BOLD signal (Boynton *et al*., 1999). While 'the jury is still out' (Georgeson and Meese, 2006) as regards the degree to which the effective transducer to explain dipper functions has a shape incorporating contrast-dependent changes in variance, this ambiguity may have an important effect on our use of the same transducer for suprathreshold magnitude ratings (J. M. Foley, personal communication). The transducer is designed to explain the results of contrast discrimination experiments, and variability in responses will be a crucial contributor to discrimination thresholds. However, this is not true for magnitude ratings, where the average rating would be expected to depend on the average neuronal response magnitude, and not on any variance (fixed or changing with contrast). Perhaps, there would be a better match between model and ratings if the magnitude and variance aspects of the effective transducer were better distinguished. If the transducer that is effective for the dipper functions does include an element of increased variance for large stimulus intensities, then this may tend to exaggerate the predicted magnitude ratings for big stimulus differences; this is consistent with the significant downward deviation from a straight line fit of the graph of ratings against model prediction in Fig. 4A.

B. H. C. Matthews (1931) proposed that the logarithmic relation between action potential firing rate and tension in a muscle receptor could be an explanation for Fechner's hopefully-universal Law of Sensation, where the ability to discriminate stimuli along simple intensity dimensions followed the rule that $\Delta I/I$ is constant. While this proposal is now seen to be too simplistic in specific detail, it was of fundamental importance in the quest to link human perceptual performance with the behaviour of individual or populations of nerve cells. Our results with complex natural visual images do, indeed, suggest that perception of difference is directly related to differences in neuronal response, but this is a population response pooled across neurons.

Acknowledgements

This research was supported by grants from the EPSRC and Dstl on the Joint Grants Scheme (EP/E037097/1 and EP/E037372/1) to D. J. Tolhurst and T. Troscianko.

M. P. S. To and P. G. Lovell were employed on those grants. M. Chirimuuta received a research studentship from the MRC. P.-Y. Chua received a studentship from the Defence Science and Technology Agency (Singapore). We thank J. A. Solomon and J. M. Foley for their challenging criticisms of this work.

References

Adrian, E. D. and Zotterman, Y. (1926). The impulses produced by sensory nerve endings, *J. Physiol. (Lond.)* **61**, 151–171.

Albrecht, D. G. and Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function, *J. Neurophysiol.* **48**, 217–237.

Barlow, H. B. and Levick, W. R. (1969). Three factors limiting the reliable detection of light by retinal ganglion cells of the cat, *J. Physiol. (Lond.)* **200**, 1–24.

Biondini, A. R. and De Martelli, M. L. F. (1985). Suprathreshold contrast perception at different luminance levels, *Vision Research* **25**, 1–9.

Blakemore, C. and Tobin, A. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex, *Exper. Brain Res.* **15**, 439–440.

Bonds, A. B. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex, *Vis. Neurosci.* **2**, 41–55.

Borg, G., Diamant, H., Strom, L. and Zotterman, Y. (1966). The relation between neural and perceptual intensity: a comparative study on the neural and psychophysical response to taste stimuli, *J. Physiol. (Lond.)* **192**, 13–20.

Boynton, G. M., Demb, J. B., Glover, G. H. and Heeger, D. J. (1999). Neural basis of contrast discrimination, *Vision Research* **39**, 257–269.

Brady, N. and Field, D. J. (2000). Local contrast in natural images: normalisation and coding efficiency, *Perception* **29**, 1041–1055.

Brodie, E. E. and Ross, H. E. (1984). Sensorimotor mechanisms in weight discrimination, *Percept. Psychophys.* **36**, 477–481.

Cannon, M. W. (1979). Contrast sensation: a linear function of stimulus contrast, *Vision Research* **19**, 1045–1052.

Cavanaugh, J. R., Bair, W. and Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons, *J. Neurophysiol.* **88**, 2530–2546.

Chirimuuta, M. and Tolhurst, D. J. (2005a). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research* **45**, 2943–2959.

Chirimuuta, M. and Tolhurst, D. J. (2005b). Accuracy of identification of grating contrast by human observers: Bayesian models of V1 contrast processing show correspondence between discrimination and identification performance, *Vision Research* **45**, 2960–2971.

Clatworthy, P. L., Chirimuuta, M., Lauritzen, J. S. and Tolhurst, D. J. (2003). Coding of the contrasts in natural images by populations of neurons in primary visual cortex (V1), *Vision Research* **43**, 1983–2001.

Daly, S. (1993). The visible differences predictor: an algorithm for the assessment of image fidelity, in: *Digital Images and Human Vision*, A. B. Watson (Ed.), pp. 179–206. MIT Press, Cambridge, Mass., USA.

Dean, A. F. (1981). The variability of discharge of simple cells in the cat striate cortex, *Exper. Brain Res.* **44**, 437–440.

Enroth-Cugell, C. and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat, *J. Physiol. (Lond.)* **187**, 517–552.

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model, *J. Optic. Soc. Amer. A, Optic. Image Sci. Vis.* **11**, 1710–1719.

Foley, J. M., Varadharajan, S., Koh, C. C. and Farias, M. C. (2007). Detection of Gabor patlerns of different sizes, shapes, phases and eccentricities, *Vision Research* **47**, 85–107.

Geisler, W. S. and Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: detection, discrimination and identification, *Vision Neurosci.* **14**, 897–919.

Georgeson, M. A. and Meese, T. S. (2006). Fixed or variable noise in contrast discrimination? The jury's still out, *Vision Research* **46**, 4294–4303.

Gescheider, G. A. (1997). *Psychophysics — The Fundamentals*. Lawrence Erlbaum Associates, USA.

Goris, R. L. T., Wichmann, F. A. and Henning, G. B. (2009). A neurophysiologically plausible population code model for human contrast discrimination, *J. Vision* **9** (15), 1–22.

Gottesman, J., Rubin, G. S. and Legge, G. E. (1981). A power law for perceived contrast in human vision, *Vision Research* **21**, 791–799.

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley, Chichester.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex, *Vis. Neurosci.* **9**, 181–197.

Heeger, D. J., Huk, A. C., Geisler, W. S. and Albrecht, D. G. (2000). Spikes *versus* BOLD: what does neuroimaging tell us about neuronal activity? *Nature Neurosci.* **3**, 631–633.

Kontsevich, L. L., Chen, C. C. and Tyler, C. W. (2002). Separating the effects of response nonlinearities and internal noise psychophysically, *Vision Research* **42**, 1771–1784.

Lauritzen, J. S. and Tolhurst, D. J. (2005). Contrast constancy in natural scenes in shadow or direct light — a proposed role for contrast-normalisation (non-specific suppression) in visual cortex, *Network, Comput. Neur. Syst.* **16**, 151–173.

Legge, G. E. and Foley, J. M. (1980). Contrast masking in human vision, *J. Optic. Soc. Amer. A, Optic. Image Sci. Vis.* **70**, 1456–1471.

Lovell, P. G., Párraga, C. A., Ripamonti, C., Troscianko, T. and Tolhurst, D. J. (2006). Evaluation of a multi-scale color model for visual difference prediction, *ACM Trans. Appl. Percept.* **3**, 155–178.

Lubin, J. (1995). A visual discrimination model for imaging system design and evaluation, in: *Vision Models for Target Detection and Recognition*, E. Peli (Ed.), pp. 245–283. World Scientific, Singapore.

Maffei, L. and Fiorentini, A. (1976). The unresponsive regions of visual cortical receptive fields, *Vision Research* **16**, 1131–1139.

Matthews, B. H. C. (1931). The response of a single end organ, *J. Physiol. (Lond.)* **71**, 64–110.

Matthews, P. B. C. (1972). *Mammalian Muscle Receptors and Their Central Actions*. Hodder and Stoughton, London.

Matthews, P. B. C. and Stein, R. B. (1969). The regularity of primary and secondary muscle spindle afferent discharges, *J. Physiol. (Lond.)* **202**, 59–82.

Meese, T. S. (2004). Area summation and masking, *J. Vision* **4**, 930–943, http://journalofvision.org/4/10/8/, doi:10.1167/4.10.8

Mullen, K. T. (1985). The contrast sensitivity of human color vision to red–green and blue–yellow chromatic gratings, *J. Physiol. (Lond.)* **359**, 381–400.

Mullen, K. T. and Kingdom, F. A. (2002). Differential distributions of red–green and blue–yellow cone opponency across the visual field, *Vis. Neurosci.* **19**, 109–118.

Murray, D. J. and Ross, H. E. (1988). E. H. Weber and Fechner's psychophysics, in: *Passauer Schriften zur Psychologiegeschichte Nr. 6, G. T. Fechner and Psychology*, J. Brozeck and H. Gundlach (Eds), pp. 79–86. Passavia Universitatsverlag, Passau, Germany.

Parker, A. J. and Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception, *Ann. Rev. Neurosci.* **21**, 227–277.

Párraga, C. A., Troscianko, T. and Tolhurst, D. J. (2005). The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model, *Vision Research* **45**, 3145–3168.

Peli, E. (1990). Contrast in complex images, *J. Optic. Soc. Amer. A, Optic. Image Sci. Vis.* **7**, 2032–2040.

Peli, E., Yang, J. A., Goldstein, R. and Reeves, A. (1991). Effect of luminance on suprathreshold contrast perception, *J. Optic. Soc. Amer. A, Optic. Image Sci. Vis.* **8**, 1352–1359.

Pointer, J. S. and Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: with emphasis on the low spatial frequency range, *Vision Research* **29**, 1133–1151.

Robson, J. G. and Graham, N. V. (1981). Probability summation and regional variation in contrast sensitivity across the visual field, *Vision Research* **21**, 409–418.

Rohaly, A. M., Ahumada, A. J. and Watson, A. B. (1997). Object detection in natural backgrounds predicted by discrimination performance and models, *Vision Research* **37**, 3225–3235.

Ross, H. E. (1995). Weber then and now, *Perception* **24**, 599–603.

Sclar, G., Maunsell, J. H. and Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey, *Vision Research* **30**, 1–10.

Simons, D. J. and Rensink, R. A. (2005). Change blindness: past, present, and future, *Trends Cognit. Sci.* **9**, 16–20.

Snowden, R. J., Treue, S. and Andersen, R. A. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns, *Exper. Brain Res.* **88**, 389–400.

Solomon, J. A. (2009). The history of dipper functions, *Attent. Percept. Psychophys.* **71**, 435–443.

Stevens, J. J. (1961). To honor Fechner and repeal his Law, *Science NY* **133**, 80–86.

Tadmor, Y. and Tolhurst, D. J. (2000). Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes, *Vision Research* **40**, 3145–3157.

To, M. P. S., Gilchrist, I. D., Troscianko, T., Kho, J. S. B. and Tolhurst, D. J. (2009). Perception of differences in natural-image stimuli: why is peripheral viewing poorer than foveal? *ACM Trans. Appl. Percept.* **6** (26), 1–9.

To, M. P. S., Lovell, P. G., Troscianko, T. and Tolhurst, D. J. (2008). Summation of perceptual cues in natural visual scenes, *Proc. Royal Soc. Lond. B. Biol. Sci.* **275**, 2299–2308.

To, M. P. S., Lovell, P. G., Troscianko, T. and Tolhurst, D. J. (2010). Perception of suprathreshold naturalistic changes in colored natural images, *J. Vision* **10**, 1–22.

Tolhurst, D. J. (1989). The amount of information transmitted about contrast by neurones in the cat's visual cortex, *Vis. Neurosci.* **2**, 409–413.

Tolhurst, D. J., Movshon, J. A. and Dean, A. F. (1983). The statistical reliability of signals in single neurones in cat and monkey visual cortex, *Vision Research* **23**, 775–785.

Tolhurst, D. J., Movshon, J. A. and Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast, *Exper. Brain Res.* **41**, 414–419.

Tolhurst, D. J., Smyth, D. and Thompson, I. D. (2009). The sparseness of neuronal responses in ferret primary visual cortex, *J. Neurosci.* **29**, 2355–2370.

Tolhurst, D. J. and Thompson, I. D. (1981). On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat, *Proc. Royal Soc. London B Biol. Sci.* **213**, 183–199.

Troy, J. B. and Enroth-Cugell, C. (1993). X and Y ganglion cells inform the cat's brain about contrast in the retinal image, *Exper. Brain Res.* **93**, 383–390.

Vogels, R., Spileers, W. and Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey, *Exper. Brain Res.* **77**, 432–436.

Watson, A. B. (1987). Efficiency of a model human image code, *J. Optic. Soc. Amer. A, Optic. Image Sci. Vis.* **4**, 2401–2417.

Watson, A. B. and Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast, *J. Vision* **5**, 717–740, doi: 10.1167/5.9.6, http://www.journalofvision.org/5/9/6/

Watson, A. B. and Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking, *J. Optic. Soc. Amer. A, Optic. Image Sci. Vis.* **14**, 2379–2391.

Werner, G. and Mountcastle, V. B. (1965). Neural activity in mechanoreceptive cutaneous afferents: stimulus-response relations, Weber functions, and information transmission, *J. Neurophysiol.* **28**, 359–397.

Wiener, M. C., Oram, M. W., Liu, Z. and Richmond, B. J. (2001). Consistency of encoding in monkey visual cortex, *J. Neurosci.* **21**, 8210–8221.

Yu, H. H., Verma, R., Yang, Y., Tibballs, H. A., Lui, L. L., Reser, D. H. and Rosa, M. G. P. (2010). Spatial and temporal frequency tuning in striate cortex: functional uniformity and specializations related to receptive field eccentricity, *Eur. J. Neurosci.* **31**, 1043–1062.