Available online at www.sciencedirect.com
SCIENCE DIRECT
Neural Networks 17 (2004) 695–705
Neural Networks
www.elsevier.com/locate/neunet

2004 Special Issue

Perceptual grouping and the interactions between visual cortical areas

Scott O. Murray*, Paul Schrater, Daniel Kersten

Department of Psychology, University of Minnesota, 75 E. River Road, Minneapolis, MN 55455, USA

Received 31 March 2004; accepted 31 March 2004

Abstract

Visual perception involves the grouping of individual elements into coherent patterns, such as object representations, that reduce the descriptive complexity of a visual scene. The computational and physiological bases of this perceptual remain poorly understood. We discuss recent fMRI evidence from our laboratory where we measured activity in a higher object processing area (LOC), and in primary visual cortex (V1) in response to visual elements that were either grouped into objects or randomly arranged. We observed significant activity increases in the LOC and concurrent reductions of activity in V1 when elements formed coherent shapes, suggesting that activity in early visual areas is reduced as a result of grouping processes performed in higher areas. In light of these results we review related empirical findings of context-dependent changes in activity, recent neurophysiology research related to cortical feedback, and computational models that incorporate feedback operations. We suggest that feedback from high-level visual areas reduces activity in lower areas in order to simplify the description of a visual image—consistent with both predictive coding models of perception and probabilistic notions of 'explaining away.'

Keywords: Feedback; Grouping; Predictive coding; Probabilistic models; Bayesian

1. Introduction

One of the extraordinary capabilities of the human visual system is its ability to rapidly select and group related elements in a complex visual scene. This capability serves to bring together information likely to belong to a common cause, such as the same contour, surface or object. Grouping also reflects a general function of cognitive systems in that it greatly simplifies the description by exploiting redundancy in the input pattern (Barlow, 1959). For example, the image of a set of parallel lines can be succinctly described as a single texture pattern ('*N* repetitions of feature X') without needing to specify each element within the pattern.

These pattern-processing capabilities appear to be reflected in the activities of neurons at various stages of the visual system. For example, the response of a neuron in V1 to a single bar oriented along a receptive field's preferred axis can be suppressed by parallel bars on the two sides or enhanced if orientations differ and a collinear bar can enhance the response (Kapadia, Westheimer, & Gilbert, 2000; Knierim & van Essen, 1992). Such pattern context effects in V1 appear to be mediated by both local connections (Das & Gilbert, 1999) and by interactions with higher areas (Hupe et al., 1998).

Grouping local features that belong to an object is particularly interesting from a physiological perspective because object shape is believed to be represented in higher stages of the visual system beyond V1, so any influence of perceived shape on lower areas would require feedback. Feedback is generally thought of as a process where activity in lower areas is positively correlated with the activity occurring in higher areas. However, recent work on predictive coding models has suggested that feedback may operate to reduce activity. In these models, higher-stages of a network compete by projecting their predictions about the stimulus to lower stages, where they are then removed from incoming data. In these models, the activity of neurons in lower stages will decrease when neurons in higher stages can 'explain' a visual stimulus, but will increase when the top-down explanation is poor (Mumford, 1992; Rao & Ballard, 1999). Other mechanisms for reducing activity via feedback are also possible and are discussed below. We present results from our own research and those of others suggesting that feedback from high-level visual areas

* Corresponding author.
E-mail address: somurray@umn.edu (S.O. Murray).

0893-6080/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2004.03.010

reduces average activity in lower areas in order to simplify the description of a visual image.

2. Feedback

Feedback projections are a prominent anatomical feature of the primate visual system (Felleman & Van Essen, 1991) and recent evidence suggests they play a critical role in visual perception (Pascual-Leone & Walsh, 2001 and see Bullier, 2001; Lamme & Roelfsema, 2000 for reviews). Nevertheless, the functional significance of these connections has been open to interpretation. There are two basic possibilities: feedback may act by modifying input-driven activity in an existing active neural population by inhibition or facilitation; or feedback may recruit or prevent activity in a new population through processes like filling in or discounting.

Evidence that cortical feedback modifies existing activity is substantial and growing. For example, selectively deactivating higher visual areas reduces context effects in lower areas (Hupe et al., 1998). Hupe et al. (1998) show, using reversible inactivation of a higher-order area (monkey area V5/MT), that feedback connections serve to amplify and focus activity of neurons in lower-order areas, and that they are important in the differentiation of figure from ground. In particular, they show that feedback connections facilitate responses to objects moving within the classical receptive field and enhance suppression evoked by background stimuli in the surrounding region. Recent observations on the 'non-classical' receptive field structure in V1 by electrophysiologists also argues for modified activity via feedback (see Angelucci, Levitt, & Lund, 2002). In addition, attention has been shown to have a profound effect on activity in primary visual cortex (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Ress, Backus, & Heeger, 2000).

Feedback models that involve recruitment have been suggested for phenomena like filling-in and illusory contour formation. Mechanisms for illusory contour formation have been modeled as within-area recruitment (Peterhans & von der Heydt, 1991). However, recruitment of neural activity in a lower visual area may result when a higher level area predicts data that is not present in the input (e.g. the well-known Dalmation dog demonstration by RC James). Using single unit recording in macaque visual cortex, Lee (2002) found V1 responses to illusory contours 35 ms after V2 responded, suggesting that feedback from V2 to V1 recruits V1 neurons (Lee, 2002; Lee, Yang, Romero, & Mumford, 2002). However, there has been little evidence of an explicit neural analog of filling-in through either within or between area processes (von der Heydt, Friedman, & Zhou, 2003).

Until recently, feedback projections were often considered to be too spatially diffuse and slow to perform computations anything more complicated than non-specific excitation or inhibition-making results that suggest targeted feedback (like those in Hupe et al., 1998) difficult to explain. However, recent anatomical and physiological evidence now suggests that feedback from higher areas to lower areas is anatomically and functionally specific. Feedback connections from extra-striate cortex target the clusters of neurons that provide feedforward projections to the same extra-striate site and there is considerable convergence of visual information to single cortical columns from extra-striate feedback (Lund, Angelucci, & Bressloff, 2003). In addition, feedback connections are very fast conducting (3.5 m/s) compared to intra-areal horizontal connections (0.1 m/s) (Girard, Hupe, & Bullier, 2001). These results are important because they demonstrate an anatomical and physiological architecture capable of providing rapid, functionally specific feedback.

2.1. Inferring feedback using fMRI

Because feedback is characterized by interactions over many areas of the brain, feedback models can be tested using methods (like fMRI) that can monitor activity in many areas of the brain simultaneously. The simplest approach is to compare the measured covariation between cortical areas to those predicted by different feedback models. However, this approach is incomplete. Feedback theories postulate causal interactions between neural activity in different areas, while measured patterns of covariation between any two areas can be non-causal—for instance both driven by a third area or fed by a common artery. Friston, Jezzard, and Turner (1994) call this distinction functional versus effective connectivity, where functional connectivity specifies the interactions between areas that can occur by any means, while effective connectivity is that subset of interactions that are causal. This kind of effect plagues most experimental designs, in which the functional role of different areas is assigned on the basis of a change in response to different kinds of stimuli. The problem is that stimuli typically differ in many ways other than just the dimensions of interest (like object category). In these cases, differentiating changes in activity that result from stimulus dimensions of interest, incidental stimulus differences, task, and interactions between brain areas is challenging.

Though distinguishing causal from non-causal interactions is notoriously difficult, causal and non-causal interactions behave differently both in terms of their timing predictions (e.g. 'A causes B' ⇒ 'A precedes B') and more fundamentally how they behave when the values of the interacting variables are manipulated (Pearl, 2000). Unfortunately, feedback effects probably occur on the order of tens of milliseconds, which is well beneath the current temporal resolution of fMRI and can even be difficult to infer based on the timing of direct physiological recordings.

Two basic strategies have been used to try to infer effective connectivity. One approach uses statistical methods like structural equation modeling to test causal

versus non-causal models via goodness-of-fit. Buchel, Coull, and Friston (1999) use this approach to infer effective connectivity changes between posterior and inferior parietal cortex during the learning of an object localization task. The other basic strategy is to try to experimentally manipulate the activity in one cortical area without directly affecting others—and looking for activity changes in other areas that result. We have used this strategy in a previous study (Murray, Kersten, Olshausen, Schrater, & Woods, 2002) which we describe in Section 3.

Finally, to use fMRI data to inform neural theories of feedback requires an understanding of the relationship between the BOLD signal and underlying neural activity. This understanding is important for any interpretation of fMRI results, but there are unique considerations for models of feedback. The BOLD signal is likely to represent changes in both non-spiking inputs to an area (dendritic currents) and spiking output activity (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Logothetis & Wandell, 2004). Though the BOLD signal in response to relatively simple stimulus parameters (e.g. motion coherence and contrast) appears to directly reflect spike rate (Boynton, Demb, Glover, & Heeger, 1999; Rees, Friston, & Koch, 2000), more complex stimulus or cognitive manipulations that involve feedback may be more difficult to interpret.

3. Experimental results

To examine the possible role of feedback during object perception, we conducted a series of fMRI experiments (Murray et al., 2002) using stimuli with features that could either be perceived as ungrouped elements or perceived as being grouped into a single perceptual 'explanation'—specifically, a single shape or object. Our first experiment used random-dot structure-from-motion stimuli. In one condition ('SFM'), random dot patterns were projected onto the surfaces of simple geometric shapes (e.g. cube, cylinder, etc.), and the shapes rotated about a single three-dimensional (3D) axis. The shapes were perceived as rigidly rotating about a single axis. In a second condition ('velocity-scrambled'), each dot's velocity from the SFM stimulus was randomly reassigned to a different dot. The result was a stimulus with all the same velocities but that lacked a coherent shape percept. Instead, the dots were perceived as moving in random directions (Fig. 1A).

We observed significant reductions in activity in V1 along with significant increases in activity in the LOC during the SFM condition (Fig. 1B). That is, when the motion velocities could be grouped into a single 3D shape activity in V1 was reduced. In addition we also observed significant reductions in activity in the motion-sensitive area MT. The differences in MT were somewhat difficult to interpret because of potential stimulus confounds. For example, the SFM stimuli had higher motion opponency. That is, in a localized region of the SFM stimulus there were more motions in opposite directions, which generally has a suppressive effect on MT activity. In addition, there was higher motion coherence—the percentage of dots moving in a single direction—in the SFM stimulus which can increase MT activity. However, neither motion opponency nor motion coherence has been shown to have any effect on V1 activity. Thus, the very large reduction in activity in V1—SFM activity was approximately half of that of the velocity-scrambled activity—was particularly noteworthy.

In a second experiment we used simple line drawings in three different configurations: (a) random lines, (b) lines that

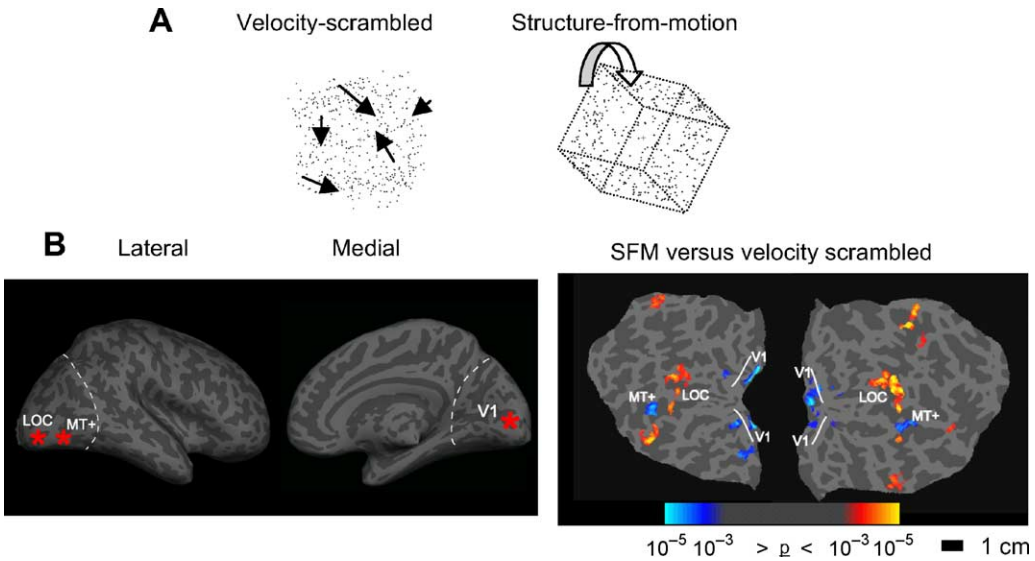

Fig. 1. (A) The two types of random dot stimuli. (B) Left, average location of the cortical areas under investigation. Right, areas of increased (red/yellow) and decreased (blue) activity comparing SFM and the velocity-scrambled control stimuli in a single subject. LOC activity increased to the SFM stimulus compared to the velocity-scrambled stimulus whereas V1 and MT + showed significant activity reductions.

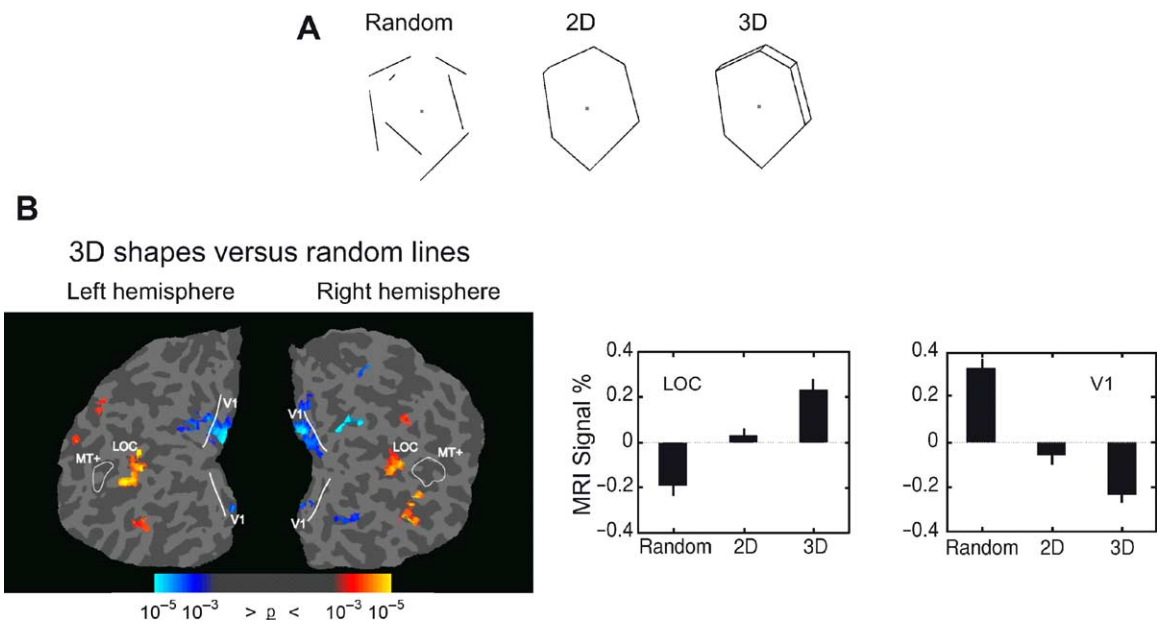

Fig. 2. (A) Examples of the three different stimulus conditions. (B) Left, areas of increased (red/yellow) and decreased (blue) activity comparing 3D figures to random lines for a representative subject on a flattened representation of occipital cortex. Right, the average percent signal change from the mean for the three conditions averaged over 6 subjects. Percent signal change is from the mean activation across all three conditions.

formed two-dimensional (2D) shapes or (c) lines that formed 3D shapes (Fig. 2A). The 2D versus 3D comparison was important because previous research had shown that the LOC codes for 3D percepts (Moore & Engel, 2001). Thus, we expected that the LOC could provide a better explanation for the lines in a 3D configuration, possibly leading to a greater reduction in activity in V1.

Consistent with this prediction, activity in the lateral occipital complex (LOC) showed a step-wise increase in response to the shape stimuli, with the 2D shapes activating the LOC significantly more than the random lines, and the 3D shapes activating LOC more than the 2D shapes. By contrast, activity in V1 showed an opposite pattern—significant reductions in response to the 2D shapes compared to the random lines and significantly less activity for 3D than 2D shapes (Fig. 2B).

Because the line-drawings occupied a restricted area of the visual field we were able to map the retinotopic specificity of the activity reductions in V1. Using a flickering checkerboard annulus that was matched to the mean eccentricity of the line drawings, we showed that the reductions in activity occurred in precisely the retinotopic location of the line-drawings. This was important because it eliminated the possibility that the reductions were a general suppressive effect of V1 or due to a non-specific hemodynamic artifact such as blood-flow steal.

Though we had performed a number of control studies showing that the results were not due to potential stimulus differences (e.g. line terminations) we wanted to make a strong case that the reductions in activity in V1 were due to the perceptual interpretation of the stimuli. A final experiment was performed that controlled for any potential stimulus based account for the effect. We used a stimulus similar to that shown in Fig. 3A, which forms a changing bi-stable percept with either grouped or ungrouped line segments. The stimulus was a line drawing of a diamond whose four corners were occluded by three vertical bars of the same color as the background. When the diamond is moved back and forth in the horizontal direction the stimulus is perceived as either a single diamond behind occluders (‘diamond’), or as separately moving line segments (‘non-diamond’). The two percepts alternated after stable intervals of several to tens of seconds and subjects indicated their perceptual state with a button press. In a recent paper by several of the authors (Murray et al., 2002), we reported significant reductions in activity in V1 when subjects perceived the diamond as compared to the non-diamond (Fig. 3D). Since publication, we have collected additional data showing that activity in the LOC increases when the diamond is perceived—consistent with both the idea that the LOC performs computations important for grouping visual features into coherent shapes and that activity in V1 is reduced as a result of activity increases in the LOC.

Our results indicate that activity is reduced in lower areas when a simpler explanation of a stimulus can be represented in a higher area. These results present a strong case for a functional role for feedback connections in the brain—it is very difficult to account for these findings with a feed-forward filtering model of vision. However, an unresolved question to answer is specifically why a *reduction* of activity is found. We address this question in the theoretical implications section below.

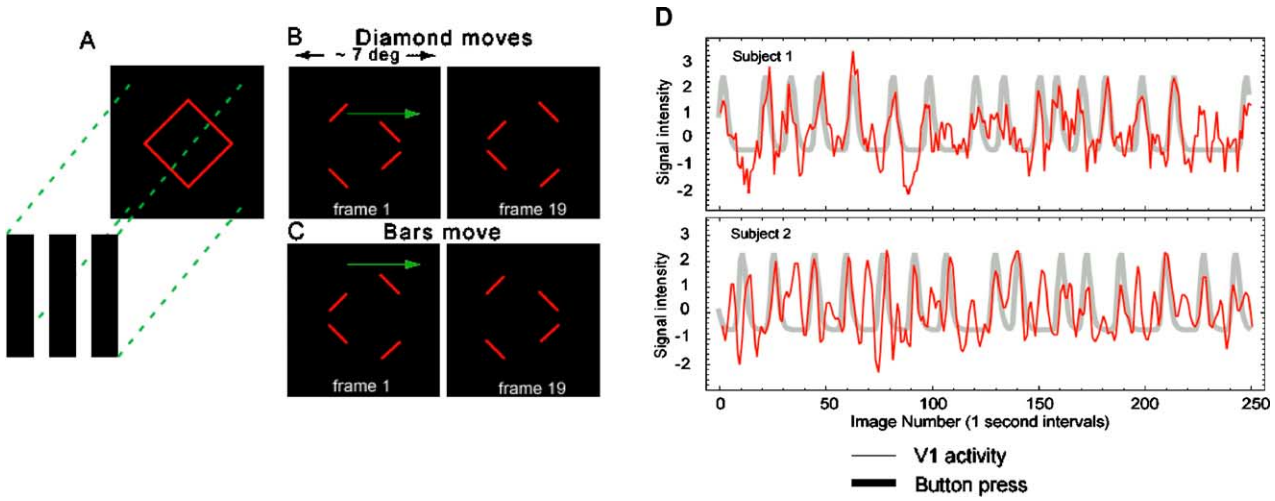

Fig. 3. (A) A red diamond was covered by three black bars that hid the four vertices. There were two stimulus conditions in which either the red diamond moved, or the three occluding black bars moved horizontally back and forth, as shown in (B) ('diamond moves') and (C) ('bars move'), respectively. The left and right columns in (B) and (C) show the first and last movie frame. The four remaining line segments could either be perceived as a rigid diamond moving horizontally or as individual line segments moving vertically. (D) Activity time course of voxels in V1 (thin red lines) and the superimposed simulated response (thick gray lines) for two subjects. In all scans for both subjects, V1 activity was an accurate predictor of subjects' perceptual state—as activity increases were associated with the non-diamond (i.e. 'ungrouped') percept.

3.1. Relation to other findings

We observed significant reductions in activity in V1 with three very different stimuli: structure-from-motion, line-drawing shapes, and the translating diamond. In all three cases the reductions in activity occurred whenever the visual features were grouped into a single shape or object—that is, whenever 'many became one'. Feedback from higher areas appears to be the primary contributor to the context effects observed in the current experiments. In our studies, the context—global shape—is a feature that can only be represented in higher stages of the visual system.

There are a variety of other experimental findings that are consistent with our results. For example, our results help to explain differences in V1 activity obtained incidentally in previous experiments. A recent study examining the effects of progressive image scrambling on LOC activity also showed significant changes in V1 activity (Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001). Specifically, it was observed that progressive decreases in LOC activity were accompanied by progressive increases in V1 activity, as images were changed from ordered to scrambled. The interpretation offered by the authors was that V1 has greater sensitivity to local image features. However, local image features were identical across the different levels of scrambling. An alternative explanation, consistent with the results of our current study, is that the changes in V1 reflect the level of perceived grouping—as higher visual areas are able to group local image features into coherent objects the need for lower areas to signal their presence is reduced.

Previous fMRI experiments that have examined depth-related shape perception have yielded similar patterns of results. The perception of shape from shading depends on the orientation of the shading gradient. For example, displays composed of elements with vertically oriented shading gradients of opposite polarity produce a strong and stable percept of 'concave' and 'convex' elements. If the shading gradients are rotated 90°, the depth percept is reduced and appears much more ambiguous. Using such stimuli, Humphrey et al. (1997) found significantly less activation in area V1 when subjects viewed displays that led to strong and stable depth percepts than when they viewed displays that led to weak and unstable depth percepts, consistent with our finding of a reduction in activity in V1 with a 3D shape percept.

If the sorts of interactions we observed between the LOC and V1 reveal a general computational principal, they should be found in other stages of the visual hierarchy. A recent fMRI experiment examined the effects of occlusion on LOC activity (Lerner, Hendler, & Malach, 2002). Subjects were presented with three types of images: (1) whole line drawings of animals ('whole'); (2) the same shapes, occluded by parallel stripes which occupied nearly half of the surface area of the images ('grid'); and (3) the same stripes, 'scrambled' so that the relative position of the regions between the stripes was changed while the local feature structure remained intact. Unlike other experiments that have used scrambled stimuli the image fragments were relatively large and contained complex feature combinations—feature combinations that have been hypothesized to be represented in intermediate visual areas such as V4. The fMRI results showed significantly higher activity in the LOC for 'whole' and 'grid' images relative to the 'scrambled' images. However, there were significant reductions in activity in areas V4/V8 and Vp for the 'whole' and 'grid' images relative to the 'scrambled'

images. There were no significant changes noted in V1. This pattern of results suggests that the specific visual area with reduced activity is dependent on the image structure that is being 'explained.' In the case of simple, low-level stimuli such as motion velocities or oriented line segments, the reductions occur in V1. With more complex stimuli, such as curved boundaries, t-junctions, and occluded surfaces, the interactions appear to occur at a higher level of the visual hierarchy.

The observed decreases in V1 activity in our experiments and others may at first appear inconsistent with the results of previous contextual modulation studies which manipulate the information (e.g. texture differences) that distinguish figure from background. For example, in previous studies a change in context that signals the presence of figure against a background resulted in increased V1 activity (Knierim & van Essen, 1992; Lamme, 1995; Zipser, Lamme, & Schiller, 1996). These observations have led some to suggest that context effects—in particular increases in activity in V1—are an index of perceptual saliency (Lamme & Roelfsema, 2000). However, one could argue that in these previous studies figure perception involved 'ungrouping', that is separating a texture patch (or line) from the background (or surrounding elements). When line segments are grouped with the background or combined into a pattern (i.e. have reduced saliency), V1 activity is reduced (Knierim & van Essen, 1992) as in the current study. When viewed from the point-of-view of 'many becoming one'—that is, achieving a simpler explanation—these figure-from-ground experiments are very consistent with our findings. Importantly, the changes in activity that have been reported in V1 for these contextual effects have occurred relatively late, suggesting the origin for the effect may be feedback from higher visual areas.

3.2. Theoretical implications

Though our results and those of others strongly suggest a role for feedback, it does not answer the question of why feedback would necessarily reduce activity. From a computational perspective, the results are consistent with two alternative accounts related to feedback modifying activity in lower visual areas. First, recent computational models of predictive coding (Mumford, 1992; Rao & Ballard, 1999)—where higher areas are attempting to actively 'explain' activity patterns in lower areas—suggest that the effect of feedback projections may be to reduce activity in lower areas. These models posit a subtractive comparison between hypotheses generated in higher areas and incoming sensory input in lower areas, with the residual from the subtraction operation passed along as 'neural activity'. Thus, reduced activity occurs when the predictions of higher-level areas match incoming sensory information (Fig. 4A).

Predictive coding models have strong intuitive appeal—why bother signaling what you already know? (Koch & Poggio, 1999). The reduced activity that would result from such a process would also have substantial biological benefits. There are clear efficiency constraints placed on the visual system—both because of inherent capacity limitations in neural pathways and because spikes are metabolically expensive (Lennie, 2003). The visual system would do well to use a representational strategy that maximizes biologically efficiency by utilizing a code that minimizes spike rate. There are, however, many ways to satisfy this efficiency constraint leading us to consider other potential mechanisms for the reduced activity we have observed in V1.

An alternative to predictive coding is that feedback may serve to sharpen activity in lower areas (Fig. 4B). In this

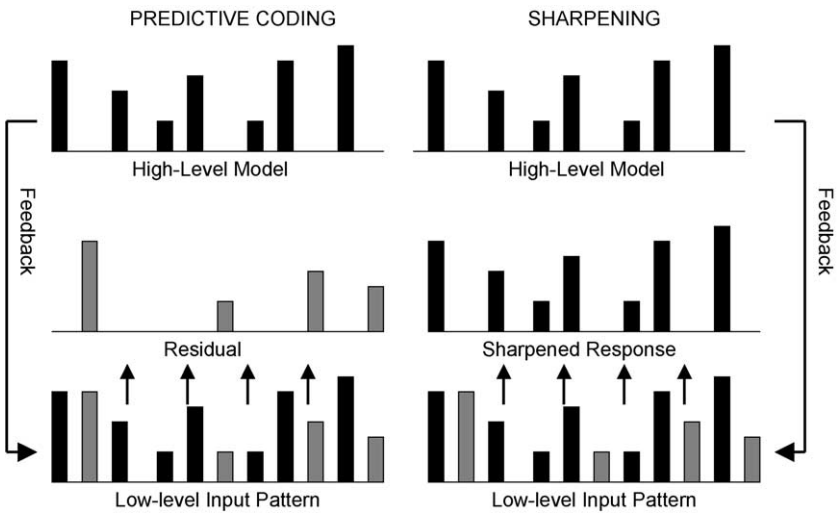

Fig. 4. Two alternative models for 'explaining away' activity with feedback from a high-level area to a low-level area. In predictive coding, a high-level model of the expected input is fed back and subtracted at the input level. What is sent forward is the difference between the expected value and the actual input. With sharpening (or 'sparsification'), the same high-level model is fed back but is instead used to amplify those aspects of the input that are consistent with the model and reduce all other aspects. The result, in both cases, is a reduction in activity.

view, feedback is hypothesized to increase activity in those aspects of the input that are consistent with predicted inputs and reduce all other activity. Averaged over a population of neurons, the result may be an overall reduction in activity. This idea is consistent with recent neurophysiological findings in V1. Vinje and Gallant (2002) investigated how the non-classical receptive field affects information transmission in V1 during viewing of natural movie stimuli in awake, behaving macaques. They varied the stimulus size from 1 to 4 times the diameter of the classical receptive field and showed, in addition to an overall decrease in activity with larger stimulus sizes, an increase in the information rate, information per spike, and the efficiency of information transmission. Though their data do not speak directly to the effects of feedback, the data suggest that increasing the amount of context increases the sparseness of the stimulus representation in V1 by tuning neurons to match the input.

4. Information processing functions of feedback

In Section 3, we discussed predictive coding and sharpening as possible explanations for the observed decrease in V1 activity as a function of shape perception. However, these ideas tell us little about how these mechanisms might be involved in solving the computational tasks of vision. Although the empirical basis for feedback between cortical areas is becoming increasingly well-established, understanding its role in information processing poses a major theoretical challenge. We can get some insight by comparing results from psychological, neural network, and computer vision studies. Theories of visual processing have traditionally emphasized a feedforward hierarchical structure, with earlier areas providing feature representations of increasing abstraction sequentially to higher areas (e.g. Fukushima, 1980; Mel, 1997; Riesenhuber & Poggio, 1999; Selfridge, 1958). Computational theories for feedback between cortical areas have received less attention. It has also been a long held assumption that feedforward processing is sufficient to solve the main time-critical problems faced by perception (Marr, 1982). The strictly feedforward assumption seems reasonable when diagnostic global features are sufficient for rapid and reliable categorization of abstract categories as for scene recognition (Oliva & Schyns, 2000; Torralba & Oliva, 2003). Rapid categorization for abstract object classes has been shown in humans (Thorpe, Fize, & Marlot, 1996; VanRullen & Koch, 2003), but this may be under those conditions where there is sufficient diagnostic information in simple global features (Johnson & Olshausen, 2003). Rapid feedforward processing could also be an efficient strategy when the trade-offs in the costs of errors are appropriate. For example, accuracy can be traded for speed for the prediction of a collision.

So what may be the possible computational functions of feedback? We briefly discuss two broad classes of theories: (1) those in which neural activity directly represents values of image features (e.g. luminance boundary), and (2) probabilistic models in which neural activity represents probability distributions on features (e.g. the likelihood of a particular object boundary being present in the image).

4.1. Feature representation theories

A common interpretation of feedback from higher to lower-level areas is to mediate attentional enhancement of early signals (Ress et al., 2000). Relatively diffuse feedback could enhance activity in neurons whose receptive fields lie within the attentional 'window' (Desimone & Duncan, 1995; Itti & Koch, 2001; Lee, Itti, Koch, & Braun, 1999; Stemmler, Usher, & Niebur, 1995; Usher & Niebur, 1996). Attentional influences may serve to increase gain and/or selectivity (Lee et al., 1999; Murray & Wojciulik, 2004). Although there is an increasing body of experimental results consistent with attentional modulation of early activity, the information processing functions for such top-down influences are not known. One possibility is for efficient visual search in high-dimensional spaces (Olshausen, Anderson, & Van Essen, 1993; Tsotsos, 1997; Tsotsos et al., 1995). Feedback could also serve to bind information across and between cortical areas. Adaptive resonance theory (Carpenter & Grossberg, 1987) and the interactive activation model of McClelland and Rumelhart (1981) both use feedback to enhance and bind low-level activity consistent with the global percept.

As discussed earlier, predictive coding theories of feedback provide an alternative and perhaps complementary view to attentional theories. Predictive coding theories of intelligent behavior have historical roots dating to the 1950s (MacKay, 1956). In its simplest form, predictive coding assumes a mechanism that computes the difference between an input value and a prediction of the input value. For example, lateral inhibition can be modeled as predictive coding in which intensity signals at a retinal location are compared with predictions based on the average from nearby retinal locations (Srinivasan, Laughlin, & Dubs, 1982). However, predictions may also arise from higher-cortical areas (Rao & Ballard, 1999). In either case, predictive coding can be interpreted as redundancy reduction resulting in a simplification of the image description (Barlow, 1959).

Interpreting feedback as predictive coding is straightforward if the signal is assumed to be image intensity information corrupted by additive noise. But the salient visual signals critical for survival are not explicit in the image and are better interpreted in terms of distal causes of intensity change, such as parameters related to object movement, shape and material. As such the retinal image is a complex encoding of relevant scene information, such as object shape boundary, but this encoding has 'noise' or sources of variation (e.g. effects of illumination including highlights, cast shadows) that cannot be simply modeled in

terms of linear additive intensity noise. Predictive coding thus could also involve more sophisticated built-in knowledge of image generative processes (Kersten, Mamassian, & Yuille, 2004). For example, a more sophisticated role for feedback is to allow for invariances that are difficult to account for using strictly feedforward processing (Grenander, 1996). For example, when searching for a match of a 2D input image to a stored memory of a 3D object, translation and scale may be compensated for on a forward pass, but a 'counter stream' carries back candidate model matches that allow for rotations in depth (Ullman, 1993, 1996).

Errors in prediction provide a measure of 'goodness of fit' which has potential utility for several functions including hypothesis refinement and learning. There are advantages to evaluating goodness of fit using the representation of an earlier level (Mumford, 1992). The explicit representation of an error residual at an earlier level may provide the means to pass information that 'needs explaining' forward to other cortical areas in a common format, resulting in hypothesis refinement. For an analogy in machine vision, shape from shading can be computed using a 'lambertian' module, leaving (non-lambertian) specularities as residuals to be explained by another module (Clark & Yuille, 1990). In another example, edges from an image could be sufficient to classify the image as a face, and a backward comparison with the input explains some of the edge information as common to all faces, but reveals other edges, that on further processing, provides shape information diagnostic of a given face (Cavanagh, 1991; Sinha & Poggio, 2001). Feedback between areas may also play an important role in long-term learning, for example as has been studied in a higher visual area, V4 (Rainer, Lee, & Logothetis, 2004). How prediction and error signals play a role in such learning will remain an important problem for the future.

4.2. Hierarchical organization of expertise and probabilistic models

For certain fundamental visual tasks, computer vision using strictly feed-forward architectures have been largely unsuccessful. One example is the automatic perceptual grouping and segmentation of objects given natural image input (Zhu, Zhang, & Tu, 2000). There are two relatively new ideas that may signal significant progress in the near future. The first comes from converging ideas from neurophysiological, psychological, and computational considerations, and that is to view a set of visual areas in terms of a hierarchical organization of expertise (Hochstein & Ahissar, 2002; Lee, Mumford, Romero, & Lamme, 1998; Zhu et al., 2000). The causal structure of images is rich and mechanisms for hierarchical inference may reflect aspects of the generative structure of image input (Kersten et al., 2004). For example, one of V1's domains of expertise may be the representation of fine spatial detail (Lee et al., 1998), whereas a higher cortical area, such as V2 may be representing longer contours and local occlusion information.

The second promising idea comes from probabilistic models of visual processing (Friston, 2003; Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1997; Weiss, 1997). In neural applications of these models, visual cortex is viewed as representing beliefs or probability distributions on feature values over levels of abstraction, corresponding to various visual areas. In the model proposed by Lee and Mumford (2003), the probability distributions are represented by a given cortical area are conditional on both the inputs and outputs. For example, beliefs in V1 are represented by $p(x_{V1}|x_{LGN}, x_{V2}, \ldots, x_{LOC})$, given lateral geniculate input, x_{LGN}, and $\{x_{V1}, x_{V2}, \ldots, x_{LOC}\}$ are variables that represent sets of feature hypotheses for various visual areas. If activity depends only on the immediate input and output, this can be simplified to $p(x_{V1}|x_{LGN}, x_{V2})$. The distribution depends on a feedforward and feedback terms derived by Bayes rule: $p(x_{V1}|x_{LGN}, x_{V2}) \propto p(x_{LGN}|x_{V1})p(x_{V1}|x_{V2})$. This suggests a probabilistic representation of hypotheses regarding image causes in which information about feature values and their uncertainty is explicitly represented and communicated between areas so that conditional distributions can be continually updated based on changes to conditional distributions in both earlier and higher areas. Under this assumption, theories of inter-cortical interaction will begin to resemble algorithms for belief-propagation that have been developed in several independent contexts (Kalman filters in signal processing, forward-backward algorithm for hidden Markov models, and Pearl's belief-propagation algorithm for inference, cf. Yedidia, Freeman, & Weiss, 2002).

One important consequence of such a view is that various areas represent and communicate their 'beliefs' in such a way as to avoid premature commitment. An important resulting idea relevant to the fMRI results presented above and discussed below is that multiple hypotheses, represented in a given cortical area, should be 'kept alive' given uncertainty. This idea corresponds to the implementation of a 'Bayesian principle of least commitment' (Kersten & Schrater, 2002). In other words, ungrouped perceptual conditions correspond to the maintenance of multiple hypotheses in V1. When the LOC arrives upon a single global interpretation, these multiple hypotheses in V1 are allowed to collapse, leading to an overall reduction in activity.

A major problem with visual Bayesian inference is learning, representing, and computing with probability distributions in high-dimensional spaces characteristic of images and their features. Here it is important to point out that all the beliefs need not be explicitly computed. Rather it is enough to maintain a state variable that summarizes the distribution. Friston (2003) and Rao and Ballard (1997) both suggest the brain need only represent a set of summary

statistics. For certain probability distributions (those from the exponential family, e.g. Gaussian), sufficient statistics exist that completely characterize the distribution.

Lee and Mumford (2003) propose another solution to the problem of computing with high-dimensional probabilities. They borrow from a powerful non-parametric method called particle filtering that has been very successful in computer vision applications such as tracking an object moving in clutter. Particle filtering approximates high-dimensional probability distributions using only a set of sample points or particles $\{x_1, x_2, \ldots, x_n\}$ and an attached set of weights $\{p_1, p_2, \ldots, p_n\}$ that represent the probabilities attached to each point. In their particle filtering model, a collection of active neurons (within and over areas) represent a particle. The probability of a particle is conjectured to be represented by rapidly adapting synaptic strengths or synchronous activity. Beliefs, sets of numbers that reflect the conditional probabilities, are passed forward and back between areas to update each area's distribution. In this way for example, the activity across V1 represents a conditional belief distribution. Thus, it is possible to simultaneously represent many alternative beliefs in V1 with sensory input and top-down beliefs serving to modulate the distribution.

A concept in Bayesian networks that may have particular importance when considering physiological models of visual perception is 'explaining away'. Explaining away is a phenomenon that occurs in probabilistic belief networks in which two (or more) variables influence a third variable whose value can be measured. Before measurement, the two causal variables are independent, but after measurement they become conditionally dependent. The phrase 'explaining away' arises because coupling of variables through shared evidence often arises in human reasoning, when the influences can be viewed as competing causes. 'Explaining away' is also a characteristic of perceptual inferences (Kersten & Yuille, 2003), for example when there are alternative perceptual groupings consistent with a set of identical or similar sets of local image features. Brain activity associated with perceptual shifts of interpretation of an image could be seen as a consequence of perceptual explaining away. For example, once the missing vertices in the translating diamond are explained by occlusion, the probability of a diamond interpretation goes up (Fig. 5). Resolving ambiguity regarding competing explanations may be a general explanation for reciprocal activity pattern of activity seen between low-and high-level areas, such as in V1 and LOC. Suppose that V1's expertise is in the representation of spatially and temporally local feature hypotheses, and LOCs is in the representation of global shape. In contrast to the conventional view of a strict functional hierarchy, these areas could be viewed as providing competing representations of the retinal input. We believe that there are sufficient unknowns in our understanding of primate functional neuroanatomy to prevent ruling out such a possibility. For example, corticothalamocortical 're-entry' may play an important role in interaction seen between cortical areas (Sherman & Guillery, 2002),

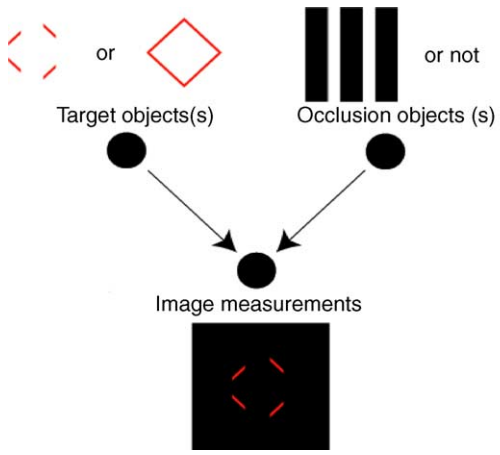

Fig. 5. Example of explaining away. Two possible object interpretations are: four line segments or one diamond object. The hypothesis of an occluder explains the missing vertices, and increases the probability of the diamond hypothesis.

5. Remaining questions

Our empirical results demonstrate that neuronal activity, even in V1, does not simply represent the signaling of features in a visual scene but is strongly influenced by high-level perceptions of object shape. Though these results, in combinations with other studies, offer a compelling example of the potential role for feedback processes in vision, there are still many unanswered questions. For example, having timing information about the relative changes in V1 and LOC is crucial to establishing that the reductions in V1 are causally linked to the increases in the LOC. Specifically, do the reductions in V1 occur after increases are observed in the LOC? It will be difficult question to answer using fMRI but may be addressable with other techniques that offer greater temporal resolution.

Understanding which model-predictive coding versus sharpening-is the correct interpretation of our findings is also going to be fundamental to understanding the computational role of feedback. While both accounts share the notion of a high-level model being fed back to 'explain away' data in a lower area, they essentially make opposite predictions about which aspects of the activity are explained (Fig. 4). In addition, because each model has clear benefits, the possibility always remains that both processes could be occurring.

Finally, understanding how our observations of reduced activity can be integrated with other concepts related to feedback (e.g. attentional competition and enhancement, resonance, learning, etc.) will be necessary. Given the extensive neural architecture in place for feedback, it is likely that feedback serves many information processing

objectives. Understanding the conditions under which these computations are used, their functions, and mechanisms will likely remain a scientific challenge for years to come.

Acknowledgements

Portions of this work were reported earlier in Murray et al. (2002) and at Human Brain Mapping (Shen, Kersten, and Ugurbil, 1999), ARVO (Kersten, Shen, Ugurbil, and Schrater, 1999) and Soc. Neurosci. (Murray Olshausen, and Woods, 2001) conferences. Supported by NIH R01 EY015261, NIH P41 RR08079, pre-doctoral NRSA MH-12791 and post-doctoral NRSA EY015342-01 (S.O.M.), NSF SBR-9631682 (D.K.), NIH MH-57921 (B.A.O.), NIH MH-41544 and VA Research Service (D.L.W.). We thank Peter Battaglia for helpful comments.

References

Angelucci, A., Levitt, J. B., & Lund, J. S. (2002). Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1. *Progress in Brain Research*, *136*, 373–388.

Barlow, H. B (1959). *Sensory mechanisms, the reduction of redundancy, and intelligence*. Paper presented at the Proceedings of the symposium on the mechanization of thought processes, National Physical Laboratory.

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research*, *39*(2), 257–269.

Buchel, C., Coull, J. T., & Friston, K. J. (1999). The predictive value of changes in effective connectivity for human learning. *Science*, *283*(5407), 1538–1541.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research. Brain Research Reviews*, *36*(2/3), 96–107.

Carpenter, G. A., & Grossberg, S. (1987). ART2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, *26*, 4919–4930.

Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (Ed.), *Representations of vision: Trends and tacit assumptions in vision research* (pp. 295–304). Cambridge, UK: Cambridge University Press.

Clark, J. J., & Yuille, A. L (1990). *Shape from shading via the fusion of specular and Lambertian image components*. Paper presented at the Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City.

Das, A., & Gilbert, C. D. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*, *399*(6737), 655–661.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352.

Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping*, *1*, 153–171.

Fukushima, K. (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202.

Girard, P., Hupe, J. M., & Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *Journal of Neurophysiology*, *85*(3), 1328–1331.

Grenander, U. (1996). *Elements of pattern theory*. Baltimore, MD: Johns Hopkins University Press.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Humphrey, G. K., Goodale, M. A., Bowen, C. V., Gati, J. S., Vilis, T., Rutt, B. K., Menon, R.S. (1997). Differences in perceived shape from shading correlate with activity in early visual areas. *Current Biology*, *7*(2), 144–147.

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience*, *2*(3), 194–203.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7), 499–512.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, *84*(4), 2048–2062.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Kersten, D, & Schrater, P. W. (2002). Pattern inference theory: A probabilistic approach to vision. In R. Mausfeld, & D. Heyer (Eds.), *Perception and the physical world*. Chichester: Wiley.

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, *13*(2), 1–9.

Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*(4), 961–980.

Koch, C., & Poggio, T. (1999). Predicting the visual world: Silence is golden [news; comment]. *Nature Neuroscience*, *2*(1), 9–10.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*(2), 1605–1615.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, *23*(11), 571–579.

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, *2*(4), 375–381.

Lee, T. S. (2002). The nature of illusory contour computation. *Neuron*, *33*(5), 667–668.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A. Optics, Image Science and Vision*, *20*(7), 1434–1448.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, *38*(15/16), 2429–2454.

Lee, T. S., Yang, C. F., Romero, R. D., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature Neuroscience*, *5*(6), 589–597.

Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*(6), 493–497.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, *11*(4), 287–297.

Lerner, Y., Hendler, T., & Malach, R. (2002). Object-completion effects in the human lateral occipital complex. *Cerebral Cortex*, *12*(2), 163–177.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, *66*, 735–769.

Lund, J. S., Angelucci, A., & Bressloff, P. C. (2003). Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cerebral Cortex*, *13*(1), 15–24.

MacKay, D. M. (1956). The epistemological problem for automata. In C. E. Shannon, & J. McCarthy (Eds.), *Automata studies* (pp. 235–250). Princeton, NJ: Princeton University Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: Freeman.

McClelland, J. L., & Rumelhart, D. L. (1981). An interactive model of context effects in letter perception. Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

Mel, B. W. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, *9*(4), 777–804.

Moore, C., & Engel, S. A. (2001). Neural response to perception of volume in the lateral occipital complex. *Neuron*, *29*(1), 277–286.

Mumford, D. (1992). On the computational architecture of the neo-cortex: II. The role of the cortico-cortical loops. *Biological Cybernetics*, *66*, 241–251.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 15164–15169.

Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nature Neuroscience*, *7*(1), 70–74.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*(2), 176–210.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, *13*(11), 4700–4719.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*(5516), 510–512.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Peterhans, E., & von der Heydt, R. (1991). Subjective contours—Bridging the gap between psychophysics and physiology. *Trends in Neuroscience*, *14*(3), 112–119.

Rainer, G., Lee, H., & Logothetis, N. K. (2004). The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biology*, *2*(2), E44.

Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, *9*(4), 721–763.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects [see comments]. *Nature Neuroscience*, *2*(1), 79–87.

Rees, G., Friston, K., & Koch, C. (2000). A direct quantitative relationship between the functional properties of human and macaque V5. *Nature Neuroscience*, *3*(7), 716–723.

Ress, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nature Neuroscience*, *3*(9), 940–945.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025.

Selfridge, O.G (1958). *Pandemonium: A paradigm for learning*. Paper presented at the Mechanisation of Thought Processes, National Physical Laboratory.

Sherman, S. M., & Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1428), 1695–1708.

Sinha, P., & Poggio, T. (2001). High-level learning of early perceptual tasks. In M. Fahle (Ed.), *Perceptual learning*. Cambridge, MA: MIT Press.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B: Biological Science*, *216*(1205), 427–459.

Stemmler, M., Usher, M., & Niebur, E. (1995). Lateral interactions in primary visual cortex: A model bridging physiology and psychophysics. *Science*, *269*(5232), 1877–1880.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, *14*(3), 391–412.

Tsotsos, J. K. (1997). Limited capacity of any realizable perceptual system is a sufficient reason for attentive behavior. *Consciousness and Cognition*, *6*(2/3), 429–436.

Tsotsos, J. K., Culhane, S., Wai, W., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artifical Intelligence*, *78*(1/2), 507–547.

Ullman, S. (1993). Sequence-seeking and counter streams: A model for information processing in the cerebral cortex. In C. Koch (Ed.), *Large scale neuronal theories of the brain*. Cambridge, MA: MIT Press.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.

Usher, M., & Niebur, E. (1996). Modelling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience*, *8*(3), 305–321.

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience*, *15*(2), 209–217.

Vinje, W. E., & Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *Journal of Neuroscience*, *22*(7), 2904–2915.

von der Heydt, R., Friedman, H., & Zhou, H. (2003). Searching for the neural mechanisms of color filling-in. In L. Pessoa, & P. De Weerd (Eds.), *Filling-in: From perceptual completion to cortical reorganization* (pp. 106–127). Oxford: Oxford University Press.

Weiss, Y (1997). Interpreting images by propagating Bayesian beliefs. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), (*Vol. 9*) (pp. 908–915). *Advances in neural information processing systems*.

Yedidia, J. S., Freeman, W. T., & Weiss, Y (2002). *Constructing free energy approximations and generalized belief propagation algorithms*. MERL Technical report TR2002-35, August.

Zhu, S. -C., Zhang, R., Tu, Z (2000). *Integrating bottom-up/top-down for object recognition by data driven Markov chain Monte Carlo*. Paper presented at the Proceedings of International Conference on Computer Vision and Pattern Recognition, SC.

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, *16*(22), 7376–7389.