The role of feedback in a hierarchical model of object perception

Salvador Dura-Bernal, Thomas Wennekers and Susan L. Denham

Abstract We present a model which stems from a well-established model of object recognition, HMAX, and show how this feedforward system can include feedback, using a recently proposed architecture which reconciles biased competition and predictive coding approaches. Simulation results show successful feedforward object recognition, including cases of occluded and illusory images. Recognition is both position and size invariant. The model also provides a functional interpretation of the role of feedback connectivity in accounting for several observed effects such as enhancement, suppression and refinement of activity in lower areas. The model can qualitatively replicate responses in early visual cortex to occluded and illusory contours; and fMRI data showing that high-level object recognition reduces activity in lower areas. A Gestalt-like mechanism based on collinearity, co-orientation and good continuation principles is proposed to explain illusory contour formation which allows the system to adapt a single high-level object prototype to illusory Kanizsa figures of different sizes, shapes and positions. Overall the model provides a biophysiologically plausible interpretation, supported by current experimental evidence, of the interaction between top-down global feedback and bottom-up local evidence in the context of hierarchical object perception.

Salvador Dura-Bernal
Centre for Robotics and Neural Systems, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK. e-mail: salvador.durabernal@plymouth.ac.uk

Thomas Wennekers
Centre for Robotics and Neural Systems, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK. e-mail: thomas.wennekers@plymouth.ac.uk

Susan L. Denham
Centre for Robotics and Neural Systems, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK. e-mail: s.denham@plymouth.ac.uk

1 Introduction

Although, traditionally, models of the visual system have focused on feedforward processes, it is becoming increasingly clear these are limited in capturing the wide range of complexities involved in visual perception. Recent reviews [4, 22] suggest that approximately only 20% of the response of a V1 neuron is determined by conventional feedforward pathways, while the rest arises from horizontal and feedback connectivity. Anatomically, feedforward sensory pathways are paralleled by a greater number of top-down connections, which provide lower areas with massive feedback from higher cortical areas [6]. Feedback terminations in the primary visual cortex (V1) are functionally organized and well-suited for centre-surround interactions, and unlike horizontal connections, their spatial and temporal properties have been found to provide an explanation for extra-classical distal surround effects [1].

Experimental evidence shows that feedback originating in higher-level areas, such as V4, IT or MT with bigger and more complex receptive fields, can modify and shape V1 responses, accounting for contextual or extra-classical receptive field effects [14, 16, 20, 31, 34, 11]. Nonetheless, the role of feedback is still far from being understood, as highlighted by the apparently contradictory effects observed in these experiments, suggesting response fields based on an intricate interaction between stimuli, surrounding context, attentional priors, and previous experience [10, 7]. A notable example is observed in V1/V2 activity in response to illusory contours with no direct retinal stimulation (e.g. Kanizsa figures), as confirmed by ERP, EEG, MEG and fMRI studies [18]. The experiments show illusory contour-related activity emerging first in Lateral Occipital Cortex (LOC), then V2 and finally in V1, strongly suggesting that the response is driven by feedback [16, 19]. Another remarkable study showed that the feedback-mediated response in foveal retinotopic cortex contains information about objects presented in the periphery, far away from the fovea [35], revealing that even the fundamental concept of a receptive field might be an inappropriate and misleading way to characterize feedback.

While there is relative agreement that feedback connections play a role in integrating global and local information from different cortical regions to generate an integrated percept [2, 17], several differing (though intersecting) approaches have attempted to explain the underlying mechanisms. In predictive coding, derived from Bayesian principles, each level attempts to predict the responses of the next lower level via feedback connections, while feedforward activity carries the error signals. Making predictions at many temporal and featural scales is an effective strategy for discovering novelty, and for verifying and refining the accuracy of internal representations, which in turn allows the system to generate better predictions. Supporting experimental evidence shows the suppression of, hypothetically error signalling, neural activity which can be contextually explained by higher-levels [25, 9].

In this paper we explore the role of feedback in object perception, taking as a starting point a biologically inspired hierarchical model of object recognition [26, 28], and extending it by adding feedback connectivity based on a biased competition architecture [5]. Hence, the model not only achieves successful feedforward recognition invariant to position and size, but is also able to reproduce modulatory

effects of higher-level feedback on lower-level activity. Finally, we extend the model to integrate a mechanism based on lateral connectivity, which solves the conceptual barriers present in obtaining spatial precision from invariant high-level abstract object representations. This enables the model to simulate illusory contour completion, present in lower visual areas, even for Kanizsa figures of different sizes and at different positions.

2 Methods/Model

This section is organized in three parts. Firstly, the feedforward architecture and operations of the hierarchical object recognition system are described. Secondly, the model is extended to include feedback connections and Belief units, which combine bottom-up and top-down information. We use a particular type of biased competition model, in which nodes suppress the inputs instead of neighbouring nodes, and the model is therefore argued to implement predictive coding. Finally, we describe the Gestalt-based algorithm employed to generate the feedback connectivity weights from complex to simple layers. These weights are used in the biased competition model to generate the feedback response in the lower simple layer, by disambiguating top-down activity from the complex layer above.

2.1 Feedforward architecture

We start by describing the basic feedforward architecture and operations which serve as the backbone of the model, and which guide the design and imposed the necessary constraints on the feedback extension. The model attempts to reproduce activity and functionality observed along the ventral or 'what' visual pathway, comprising areas V1, V2, V4 and IT. It is based upon widely accepted basic principles, of cortical object recognition such as a hierarchical arrangement of these areas, with a progressive increase in receptive field size and complexity of preferred stimuli, as well as a gradual build-up of invariance to position and scale as we go further up the hierarchy. At the lowest level we observe simple V1 neurons, with small receptive field tuned to basic oriented gratings, while at the other end of the spectrum evidence shows IT neurons associated with complex invariant object-like representations [13]. The architecture of the model is arranged in 3 levels, roughly representing areas V1, V2/V4 and IT; and each level is composed of 2 layers, simple and complex, stemming from the basic Hubel and Wiesel [12] proposal.

Two operations are performed in alternating layers of the hierarchy: the invariance operation, which occurs between layers of the same level (e.g. from S1 to C1); and the selectivity operation implemented between layers of different levels (e.g. from C1 to S2). Invariance is implemented applying the max function over a set of

afferents selective to the same feature, but with slightly different positions and sizes, achieving responses which are invariant to translation and scale.

An initial training stage is required to learn the S2 prototypes in an unsupervised manner. The 200 prototypes are learned by extracting patches, each one composed of 144 elements (3 x 3 positions x 4 orientations x 4 scale bands), from the C1 layer response. To ensure prototypes are statistically meaningful we used the k-means algorithm, an unsupervised clustering technique particularly well-suited to extract Radial Basis Function centres. The resulting prototypes are therefore a weighted combination of C1 invariant Gabor-like features at different orientations, positions and scales. This generates units with bigger RF sizes and with a wide variety of complex shape selectivities, as shown in the small sample in Figure 1.

The 60 S3 prototypes are learned in a supervised way using as input each of the objects in the training set (see Figure 1), and extracting a representative 6 x 6 patch from the C2 response. This patch, which constitutes an abstract representation (scale and position invariant) of each of the objects in the training set, is stored as the S3 prototype which can be used for template-matching during the recognition phase.

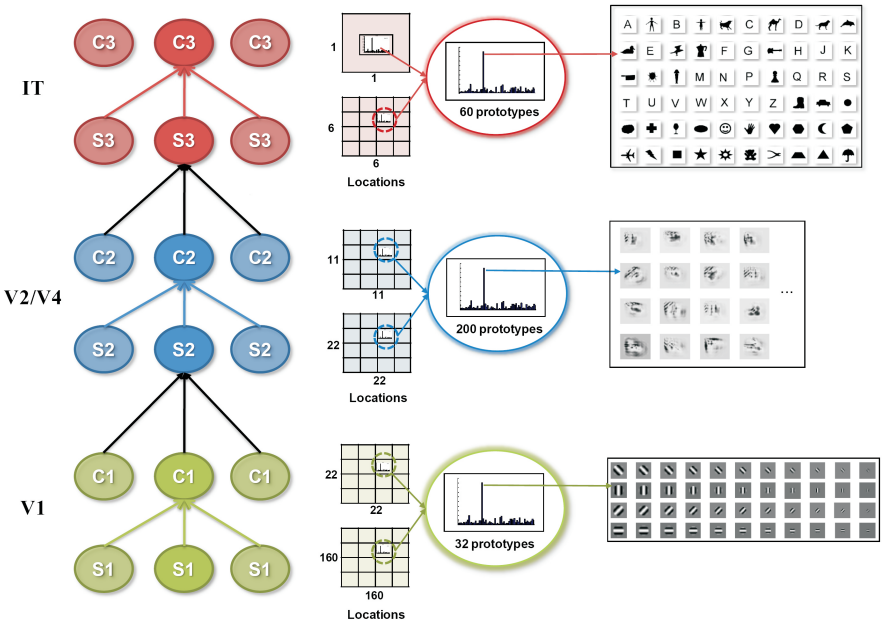

Fig. 1 Feedforward architecture. Left: Structure of units. Each location of the image is encoded by a set of units tuned to different stimuli, which increase in size and complexity along the hierarchy. The number of locations and prototypes for each layer is shown. Right: Graphical representation of prototypes at bottom level (32 Gabor filters), intermediate level (200 abstract features/object parts), top level (60 2D objects).

The implementation details for each of the layers are summarized in Table 1. The pooling over previous layer refers to the number of units used as input to the function (similar to the receptive field). The shift between units is an indication of the overlap between the receptive fields of the upper level units. The parameters of the model were based on existing HMAX implementations [26, 29, 3], to which we refer for a more detailed description. For supporting neurophysiological evidence see the Discussion section.

Table 1 Model feedforward parameters

Parameter	S1	C1	S2	C2	S3	C3
Operation performed	Gabor filter	MAX	RBF[1]	MAX	RBF	MAX
Pooling area over previous layer	7x7... 37x37	14x14	3x3	3x3	6x6	6x6
Shift between units	1	8	1	2	1	1
Number of prototypes	32	16	200	200	60	60
Number of spatial locations	160x160	22x22	22x22	11x11	6x6	1

[1] Radial Basis Function

2.2 Feedback extension

The proposed extended model, shown in Figure 2, employs three different types of units at each level to encode the feedforward error signal (eS and eC units), feedback predictive signal (*fS* and *fC* units) and Belief (*B* units). This yields an architecture similar to a recent model which reconciles biased competition with predictive coding approaches [32]. The main difference is the fact that each level now consists of two layers, i.e. simple and complex. Belief units at each level maintain an active representation of the stimuli, which is multiplicatively modulated by bottom-up error and top-down prediction.

The bottom-up error at a given level is a function (f_{Fsel}) of the complex error units (eC) in the level below and is symbolically labeled λ (from likelihood). Complex error units (eC) are a function (f_{Finv}) of simple error units (eS). These, in turn, are calculated by dividing the Belief (B) at that level by the top-down prediction (π). The top-down prediction at each level is given by the simple feedback units (fS) at that level and is symbolically labeled π (from prior). Simple feedback units (fS) are a function (f_{Binv}) of complex feedback units (fC). These, in turn, are a function (f_{Bsel}) of the Belief (B) in the level above.

Note Figure 2 is a simplified schematic version of the model architecture, where each oval represents the set of units coding all the different features and locations of that layer. The number and properties of units in each layer is defined by the feedforward parameters described in the previous section.

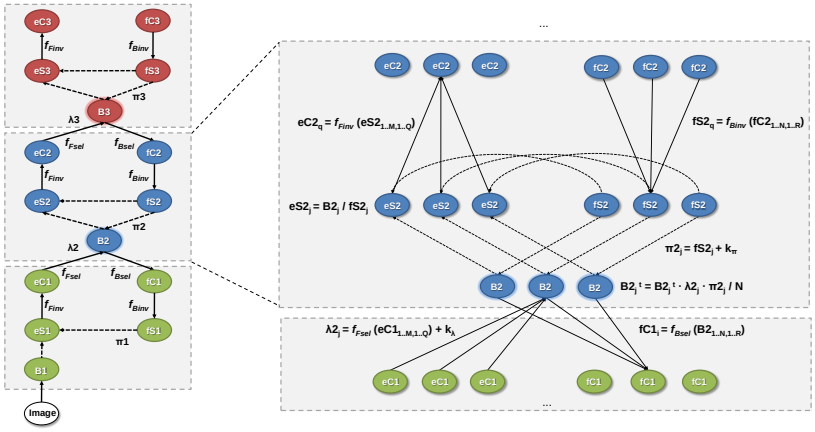

Fig. 2 Architecture of the biased competition/predictive coding implementation of the invariant hierarchical object perception model. Left: Schematic representation of the whole model (3 levels). Right: Detailed 'zoomed in' diagram the intermediate level. Belief units at each level maintain an active representation of the stimuli, which is multiplicatively modulated by bottom-up error and top-down prediction. See text for details.

The following simplified equations define the dynamics of the system presented in Figure 2, where in order to generalize the equations, the indices i, [i-1] and [i+1] are used to denote the level of the unit:

$$
\begin{aligned}
\mathrm{B}i^t &= Bi^{t-1} \cdot \lambda i^t \cdot \pi i^t / N \\
\lambda i^t &= f_{Fsel}(W^{eC[i-1]}, eC[i-1]^t) + k_\lambda \\
eCi^t &= f_{Finv}(W^{eSi}, eSi^t) \\
eSi^t &= Bi^t / fSi^t \\
\pi i^t &= fSi^t + k_\pi \\
fSi^t &= f_{Binv}(W^{fSi}, fCi^t) \\
fCi^t &= f_{Bsel}(W^{fCi}, B[i+1]^t) \quad (1)
\end{aligned}
$$

where t = time in discrete steps; k_λ = feedforward noise constant; k_π = feedback noise constant; N = normalizing constant; W^{eSi}, W^{eCi}, W^{fSi}, W^{fSi} = weights between units in different layers; f_{Fsel}, f_{Finv}, f_{Bsel}, f_{Binv} = functions linking units in different layers; B1 = Evidence = $\mathrm{W}^{eC0} \cdot$ Image.

The feedforward functions for selectivity, f_{Fsel} , and invariance, f_{Finv}, are derived from the feedforward operations in the HMAX model, namely, the Radial Basis Function (approximated as a dot-product) and MAX operations respectively. This means, during the initial time-step, when we assume there is no top-down modulation, the error units (*eS and eC*) have equivalent properties to the units in the standard feedforward model (HMAX).

To complete the definition of the model we need to specify the feedback selectivity operation, f_{Bsel}, that will link Belief units (B) with complex feedback units in the level below (eC). Taking into account that feedforward Radial Basis Function can be approximated by a weighted sum operation, the feedback function can be trivially obtained by appropriately inverting the prototype weights, and applying an analogous weighted sum operation. However, the feedback invariance function, f_{Binv}, requires some additional processing to obtain the appropriate weights. The specific algorithm developed for this purpose is described in a subsequent section. The equations for the feedforward and feedback functions are as follows:

$$\begin{aligned}
&\text{eS[i+1]} = f_{Fsel}(eCi) = \textstyle\sum_{m=1}^{M}\sum_{q=1}^{Q}(eCi_{m,q}\cdot W_{m,q})\\
&eCi = f_{Finv}(eSi) = max_{m=1}^{M}max_{q=1}^{Q}(eCi_{m,q}\cdot W_{m,q})\\
&fCi = f_{Bsel}(B[i+1]) = \textstyle\sum_{n=1}^{N}\sum_{r=1}^{R}(B[i+1]_{n,r}\cdot W_{n,r})\\
&fSi = f_{Binv}(fCi) = max_{n=1}^{N}max_{r=1}^{R}(fCi_{n,r}\cdot W_{n,r}) \quad (2)
\end{aligned}$$

By adjusting the values of k_λ and k_π in the set of equations 1 we can vary the modulatory effect that the feedforward and feedback components will have on the Belief. We assume k_π is fixed to 1, so feedback is always excitatory, while we make k_λ proportional to the ambiguity (measured as the inverse of the standard deviation) of the error signal in relation to the Belief, $k_\lambda = \sigma(B_i^t)/\sigma(\lambda_i^t)$. This means when the Belief is ambiguous and the error signal is highly informative, the modulatory effect of the error signal will be high (e.g. during the initial time step). In contrast, when the Belief has been already established and the error signal is low, its modulatory effect will be limited.

The normalizing constant N determines the overall level of enhancement/reduction of the Belief, and is calculated as $N = 1/(max(\lambda_i^t + k_\lambda)\cdot mean(\pi_i^t + k_\pi))$. This value tries to ensure relative stability in the resulting Belief, by maintaining the population activity within a certain range. However, due to the intrinsic dynamics of the model, certain values of the distribution will be enhanced while others will be reduced.

For the feedback invariance operation, we require a weight matrix which relates a complex node with each of its simple children nodes. Note that this information is not contained in the prototype weight matrix learned during the training phase, as the feedforward invariance (max) operation is non-linear, and doesn't employ a weight matrix which can be inverted. In other words, the precise detailed information is lost as one moves up the hierarchy leading to scale and size invariant representations at the top level.

Therefore to obtain the required feedback weights we propose a novel disambiguation algorithm which maps the response of one single complex unit to many simple units. To do this the algorithm uses existing feedforward responses as cues and implements extrapolation techniques based on collinearity, co-orientation and good continuation principles [15]. To illustrate this, Figure 3 shows how activity

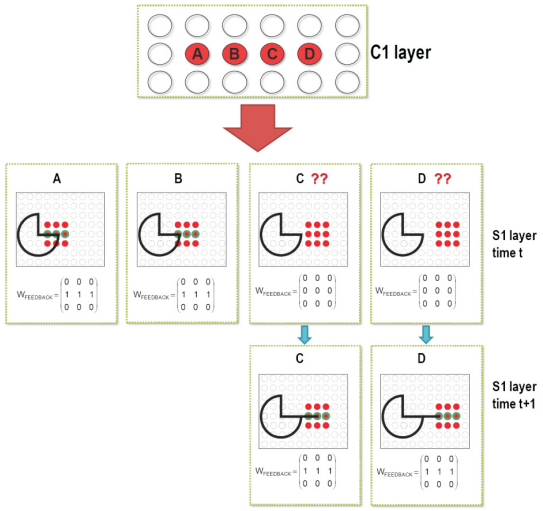

Fig. 3 Disambiguation algorithm which calculates the weight matrix for the feedback invariance function. Activity from the C1 layer (orange circles labeled A to D) feeds back to the corresponding set of S1 units (smaller orange circles). The distributed feedback is then disambiguated using local evidence precisely represented in the S1 layer (black line), and applying extrapolation methods to generate the feedback weight matrix (green circles). The resulting feedback weight matrix ($W_{FEEDBACK}$) is shown under each S1 patch. See text for details

from the C1 layer (orange circles labelled A to D) feeds back to the corresponding set of S1 units (smaller orange circles). The distributed feedback is then disambiguated by using local evidence precisely represented in the S1 layer (black line), and by applying extrapolation methods to generate the feedback weight matrix (green circles). Initially, there is not enough local evidence at t=1 to disambiguate the feedback from units C and D. Only at t=2, after the S1 contour has been extended due to the effect of units A and B, can this feedback be disambiguated based on the new local evidence. This effect is only possible due to the overlap present in the model, which means the same S1 unit can receive feedback from several C1 units. The response amplitude, the length and the orientation of the local activity must meet some minimum thresholds in order to be used as disambiguation cues.

3 Results

3.1 Feedforward recognition

The network was trained using 60 object silhouette images from which the S2 and S3 prototypes were learned. The trained network was then tested on different transformations of the same images including occluded, translated and scaled versions (Figure 4A). The experiment was repeated for a second set of 60 natural images (Figure 4B).

For the occluded test set an average of 30% of the image's black pixels are deleted using a rectangular white patch. The rectangle is placed in a position which makes

the image still identifiable to a human observer. In the translated test-set the object is moved to a new position within the available image frame of 160x160 pixels. The displacement will be near to the maximum permitted in both directions but will depend on the original object size, i.e. small objects allow for bigger displacements. Two different scale sets have been used: scale 10% where the image is scaled to either 90% or 110% of the original size and centred; and scale 20% where the image is scaled to either 80% or 120% of the original size and centred. Additionally, to test the degree of scale invariance we used an individual object at different scales (10%, 20% and 30%) (Figure 4C).

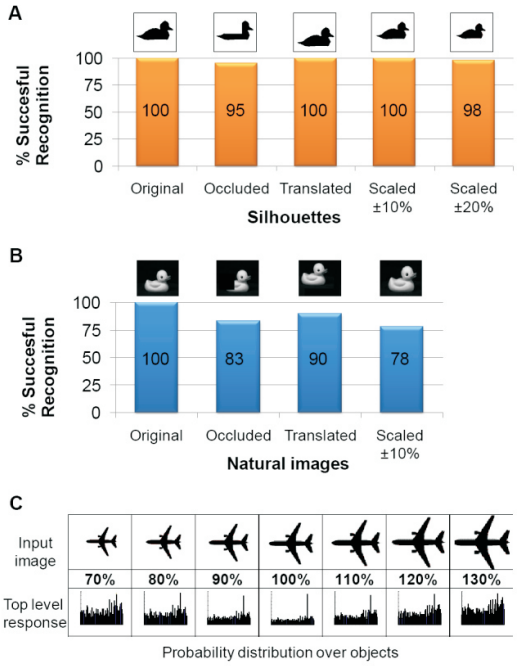

Fig. 4 Feedforward selectivity and invariance results. A) The model was trained using a dataset containing 60 silhouette objects. It was then tested using occluded, translated and scaled versions of these objects. The results, obtained during the first feedforward pass in the top layer (eC3), show the percentage of correctly categorized objects for each of the categories. An example image is shown for each of the categories. B) Analogous results for 60 natural images of objects. C) Size-invariant recognition results. The probability distribution over the 60 objects of the dataset is shown for each of the scaled versions of the same object. Although noise increases proportionally to the scaling percentage, the image is still correctly categorized in all cases.

The average percentage of successfully categorized objects was 98% for silhouettes and 88% for natural images, demonstrating the tolerance of the model to variations in object location, scale and occlusions. Correct categorization occurs when, during the first time step, the highest value of the probability distribution over objects at the eC3 layer matches the input object.

3.2 Feedback modulation

Figure 5A shows the temporal evolution of the lower level prediction c1F and s1F derived from the intermediate level Belief, B2, when using a Kanizsa illusory figure as input image. Layer c1F shows a gradual shift from the '4 pacmen' to the 'square' figure due to the effect of high-level feedback from B3 imposing its Belief (the image was recognized as a 'square' object). Subsequently, activity in s1F shows a gradual development of the illusory contour due to the disambiguation algorithm.

Figure 5B shows the model response after feedback for different versions of the Kanizsa square, including translated 10 pixels to the right, scaled 90%, scaled 110%, with rectangular symmetry, and with blurred edges. Results show how a single high-level prototype manages to complete the different distorted version of the Kanizsa square. Note the contour completion will only occur when the variations lies within the position and scale invariance range of the high-level prototype; and when there is enough local evidence to guide the extrapolation/interpolation process. The second condition is not satisfied in the figure with blurred edges.

We compared the activity in the error layers before and after feedback modulation using an object present in the training set, i.e. recognized unambiguously. Error units in layers eS1 and eC1, which code the difference between the bottom-up input (*B1*) and the high-level prediction (*B2*) showed high activity in response to a novel input image which cannot yet be predicted. This activity will then propagate upwards to B2, and again the error populations eS2 and eC2 will yield high activity as the higher level (B3) still provides no prediction. Once B2 and B3 have been updated, the high-level predictions are fed back through the feedback units and strongly reduce the response of the error units to an average value of approximately 18% of the original value (due to space limitations these results are not shown in the present reduced version of the manuscript).

Figure 5C shows an example of Belief refinement, where only a small subpopulation of units consistent with the high-level prediction is enhanced, while the rest are suppressed.

4 Discussion

The model described extends an existing feedforward hierarchical model of object recognition to include feedback connectivity and provides a framework in which both sources of information can be integrated. Feedback is implemented as a biased competition architecture consistent with predictive coding principles. The model is constrained by selectivity and invariance properties found in neurons in the different simulated cortical areas, as well as by the general principles governing hierarchical object recognition in the ventral pathway [26, 29]. Integrating top-down influences, mediated by feedback projections, with bottom-up processing had been previously pointed out as one of the main limitations and future challenges for this model [27]. Our proposed model addresses this challenge.

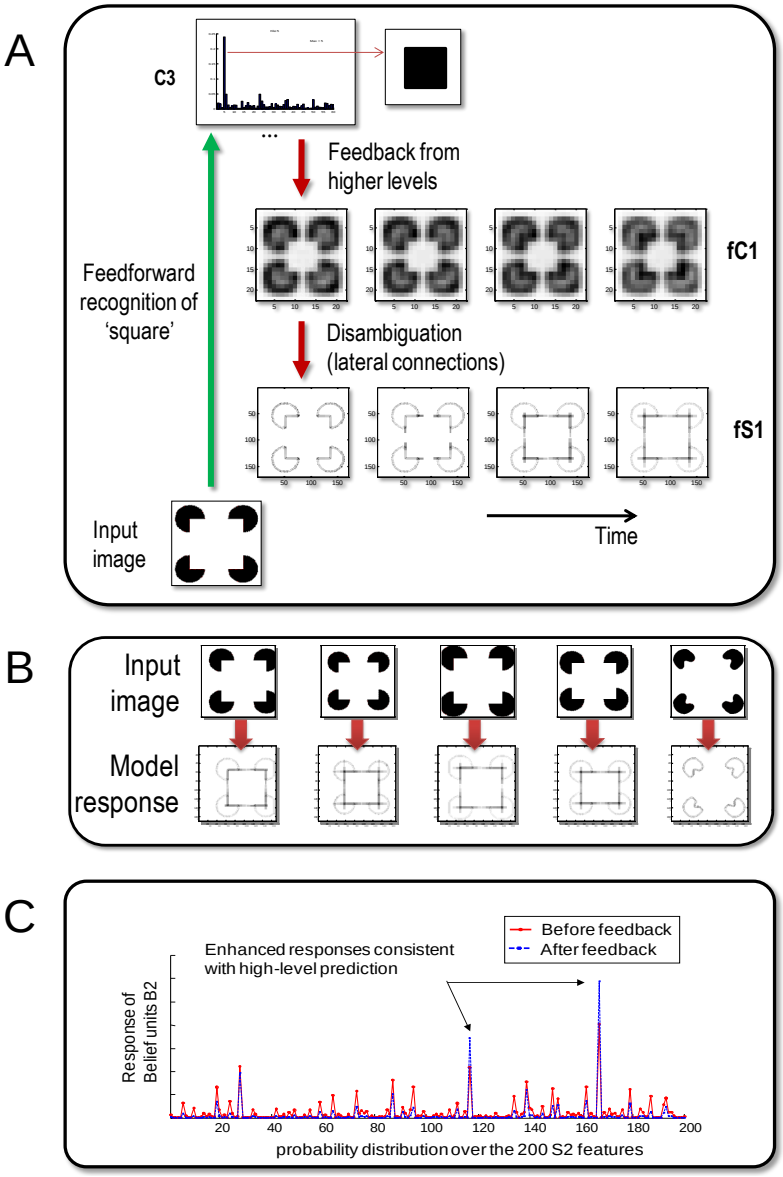

Fig. 5 A) Illusory contour completion as a result of the combination of higher-level feedback, interpolation and extrapolation, consistent with recent studies[8]. B) Feedback from single high-level prototype completing distorted version of the Kanizsa's square. See text for details. C) Response of Belief units before (red line) and after feedback (blue dotted line). Only a small subpopulation of units consistent with the high-level prediction is enhanced, while the rest are suppressed.

Overall, physiological data on simple and complex RF size, spatial frequency and orientation bandwidth are in good agreement with the model S1 and C1 tuning properties, as well as with the hypothesis of complex cells performing a MAX operation over simple cell afferents [29]. As for the upper levels, It has been shown that the S2-C2 hierarchy produces both selectivity and invariance that matches observed responses in V4 [3]. Although the implementation of top-level units S3-C3 varies between previous versions of HMAX, e.g. from Gaussian tuning [26] to Support Vector Machines [28], the overall concept is preserved and captured by the current model. Top level units present bigger RFs and are tuned to complex composite invariant features, which are consistent with the so-called view-tuned cells present in the higher levels of the ventral pathway, such as the infero-temporal cortex [13, 23, 30].

Regarding feedback, we hypothesize the proposed mechanism arises from the interaction between feedback and lateral connections in layers 2/3. Lateral connections are responsible for collinear and co-oriented facilitation in the proximal surround [1]. Additionally, the mechanism is in harmony with previous models describing the interaction between feedback and lateral connections: the preattentive-attentive mechanism resolving perceptual grouping [24], models of spatial integration and perceptual grouping [21].

Results show the response in eS1 and eC1 is strongly reduced indicating that the level above captures most of the information present in the input image. Analogously, the response in eS2 and eC2 is also significantly decreased as the top-down feedback from B3 manages to predict the activity in B2. The general refinement of the Belief is aimed at reducing both the lower-level error, eS1 and eC1, by achieving a more accurate reconstruction of the input; and the upper-level error, eS2 and eC2, by increasing the strength of features consistent with the high-level representation in B3.

These results reconcile previous apparently contradictory approaches by using two distinct populations: Belief units which show refinement of the response due to feedback (both enhancement of consistent features and reduction of redundant ones), supported by experimental evidence suggesting feedback enhances activity consistent with the high-level prediction [14, 11]; and the error population, which are suppressed due to feedback, consistent with evidence showing reduction of lower-level activity due to feedback, and with predictive coding approaches [25, 9].

Overall, the illusory contour emerges as a consequence of the interaction between the global contextual feedback signal and horizontal connectivity, guided by existing feedforward cues (local evidence). The model is consistent with a recent review on illusory contour formation which hypothesizes three mechanisms are responsible for the phenomenon: interpolation, extrapolation and figural feedback [8]. The model is also supported by evidence showing the illusory contour response in V2 precedes that in V1; and the finding that lesion of IT impaired a monkey's ability to see illusory contours [17]. Further substantiating proof was provided by [19] who showed illusory contour sensitivity first occurred in LOC, a high-level area in the ventral stream.

In addition, [33] tested Kanizsa squares with blurred edges and found out the response in the LOC area was similar to that of sharp edged Kanizsa squares, although psychophysical experiments demonstrated that the perceived boundary was not as sharp and less well localized. This in turn suggests contour-based processes that support the perception of illusory contours are performed in early retinotopically organized visual areas (V1, V2) where it should be possible to observe differential responses to the blurred-type stimuli. Testing a blurred Kanizsa square in the model reproduced these results (right column in Figure 5B): the higher-level recognized the image as a square, but at the lower level's the illusory contour could not be accurately reconstructed as there was not enough precise local information in the image to disambiguate the feedback.

Another important property of the dynamic disambiguation algorithm is that feedback, which emerges from an abstract invariant high-level prototype, can adapt to match object variations which lie within the invariance range of the prototype. This means one single square prototype can be shaped by lateral connections, on the basis of local evidence, to complete the illusory contours of Kanizsa squares at different locations, scales and proportions, as shown in Figure 5B). This contrasts with the blurred contours figure where, despite feedback, local cues are not precise enough to enable the completion of the illusory square.

The model can also be used to make predictions about the responses of different neural populations at each cortical level (e.g. error and representation units) and their expected pattern of activity in response to different stimuli (e.g. refinement or suppression). Further, the framework allows for extensions such as learning and plasticity during feedback, which can enhance the stored prototypes based on new data thus improving recognition, and adaptation which could lead naturally to phenomena such as sensitivity to temporal context and bistability.

References

1. Angelucci, A., Bullier, J.: Reaching beyond the classical receptive field of vi neurons: horizontal or feedback axons? Journal of Physiology-Paris **97**(2-3), 141–154 (2003)
2. Bullier, J.: Integrated model of visual processing. Brain Research Reviews **36**(2-3), 96–107 (2001)
3. Cadieu, C., Kouh, M., Pasupathy, A., Connor, C.E., Riesenhuber, M., Poggio, T.: A model of v4 shape selectivity and invariance. Journal of Neurophysiology **98**(3), 1733–1750 (2007)
4. Carandini, M., Demb, J.B., Mante, V., Tolhurst, D.J., Dan, Y., Olshausen, B.A., Gallant, J.L., Rust, N.C.: Do we know what the early visual system does? Journal of Neuroscience **25**(46), 10,577–10,597 (2005)
5. Deco, G., Rolls, E.T.: A neurodynamical cortical model of visual attention and invariant object recognition. Vision Research **44**(6), 621–642 (2004)
6. Felleman, D., Van Essen, D.: Distributed hierarchical processing in primate cerebral cortex. Cerebral Cortex **1**(1), 1–47 (1991)
7. Gilbert, C.D., Sigman, M.: Brain states: Top-down influences in sensory processing. Neuron **54**(5), 677–696 (2007)
8. Halko, M.A., Mingolla, E., Somers, D.C.: Multiple mechanisms of illusory contour perception. Journal of Vision **8**(11), 1–17 (2008)

9. Harrison, L.M., Stephan, K.E., Rees, G., Friston, K.J.: Extra-classical receptive field effects measured in striate cortex with fmri. Neuroimage **34**(3), 1199–1208 (2007)
10. Hochstein, S., Ahissar, M.: View from the top: Hierarchies and reverse hierarchies in the visual system. Neuron **36**(5), 791–804 (2002)
11. Huang, J.Y., Wang, C., Dreher, B.: The effects of reversible inactivation of postero-temporal visual cortex on neuronal activities in cat's area 17. Brain Research **1138**, 111–128 (2007)
12. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of Neurophysiology **28**, 229–289 (1965)
13. Hung, C.P., Kreiman, G., Poggio, T., Dicarlo, J.J.: Fast readout of object identity from macaque inferior temporal cortex. Science **310**(5749), 863–6 (2005)
14. Hupe, J.M., James, A.C., Girard, P., Lomber, S.G., Payne, B.R., Bullier, J.: Feedback connections act on the early part of the responses in monkey visual cortex. Journal of Neurophysiology **85**(1), 134–145 (2001)
15. Keane, B.P., Lu, H., Kellman, P.J.: Classification images reveal spatiotemporal contour interpolation. Vision Research **47**(28), 3460–3475 (2007)
16. Lee, T., Nguyen, M.: Dynamics of subjective contour formation in the early visual cortex. Proceedings of the National Academy of Sciences **98**(4), 1907–1911 (2001)
17. Lee, T.S.: Computations in the early visual cortex. Journal of Physiology-Paris **97**, 121–139 (2003)
18. Maertens, M., Pollmann, S., Hanke, M., Mildner, T., Mller, H.E.: Retinotopic activation in response to subjective contours in primary visual cortex. Frontiers in Human Neuroscience **2**(2), doi:10.3389/neuro.09.002.2008 (2008)
19. Murray, M.M., Wylie, G.R., Higgins, B.A., Javitt, D.C., Schroeder, C.E., Foxe, J.J.: The spatiotemporal dynamics of illusory contour processing: Combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. Journal of Neuroscience **22**(12), 5055–5073 (2002)
20. Murray, S.O., Schrater, P., Kersten, D.: Perceptual grouping and the interactions between visual cortical areas. Neural Networks **17**(5-6), 695–705 (2004)
21. Neumann, H., Mingolla, E.: Computational neural models of spatial integration in perceptual grouping. In: T.F.Shipley, P. Kellman (eds.) From Fragments to Objects: Grouping and Segmentation in Vision, pp. 353–400. Elsevier, Amsterdam (2001)
22. Olshausen, B., Field, D.: How close are we to understanding v1? Neural Computation **17**(8), 1665–1699 (2005)
23. Quiroga, Q., Reddy, L., Kreiman, G., Koch, C., Fried, I.: Invariant visual representation by single neurons in the human brain. Nature **435**(7045), 1102–1107 (2005)
24. Raizada, R.D.S., Grossberg, S.: Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. Cerebral Cortex **13**(1), 100–113 (2003)
25. Rao, R.P.N., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience **2**(1), 79–87 (1999)
26. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. Nature Neuroscience **2**(11), 1019–25 (1999)
27. Riesenhuber, M., Poggio, T.: Models of object recognition. Nature Neuroscience (2000)
28. Serre, T., Oliva, A., Poggio, T.: A feedforward architecture accounts for rapid categorization. Proceedings of the National Academy of Sciences **104**(15), 6424–6429 (2007)
29. Serre, T., Riesenhuber, M.: Realistic modeling of simple and complex cell tuning in the hmax model, and implications for invariant object recognition in cortex. Massachusetts Institute of Technology, Cambridge, MA. **CBCL Paper 239/AI Memo 2004-017** (2004)
30. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(3), 411–426 (2007)
31. Sillito, A.M., Cudeiro, J., Jones, H.E.: Always returning: feedback and sensory processing in visual cortex and thalamus. Trends in Neurosciences **29**(6), 307–316 (2006)
32. Spratling, M.: Reconciling predictive coding and biased competition models of cortical function. Frontiers in Computational Neuroscience **2**(4), 1–8 (2008)

33. Stanley, D.A., Rubin, N.: fmri activation in response to illusory contours and salient regions in the human lateral occipital complex. Neuron **37**(2), 323–331 (2003)
34. Sterzer, P., Haynes, J.D., Rees, G.: Primary visual cortex activation on the path of apparent motion is mediated by feedback from hmt+/v5. Neuroimage **32**(3), 1308–1316 (2006)
35. Williams, M.A., Baker, C.I., Op de Beeck, H.P., Mok Shim, W., Dang, S., Triantafyllou, C., Kanwisher, N.: Feedback of visual object information to foveal retinotopic cortex. Nature Neuroscience **11**(12), 1439–1445 (2008)