Top-down influence in early visual processing
A Bayesian perspective

Tai Sing Lee

Center for the Neural Basis of Cognition
Department of Computer Science
Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
Department of Neuroscience
University of Pittsburgh, Pittsburgh, PA 15213, U.S.A.

Address: Professor Tai Sing Lee
Rm 115, Mellon Institute, Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, U.S.A.
Tel: 412 268 1060
Fax: 412 268 5060
E-mail: tai@cnbc.cmu.edu

Running head: Top-down influence on early visual processing.

Keywords: primate electrophysiology, visual processing, feedback, saliency, contours.

Abstract

LEE, T.S. Top-down influence in early visual processing: a Bayesian perspective. PHYSIOL BEHAV 56(6) 000-000, 2002. Traditional views of visual processing suggest that early visual neurons are static spatiotemporal filters that extract local features by feedforward computation. The extracted information is then channelled through a chain of modules to successively higher visual areas for further analysis. Recording from early visual neurons in awake behaving monkeys, we revealed there are many levels of complexity in the information processing of the early visual cortex. We found that the early visual neurons not only are sensitive to features within their receptive fields but also to the global context of a visual scene, the statistics of the environment and the behavioral relevance of the visual stimuli. These findings suggest that the early visual cortex (V1 and V2) is tightly coupled and highly interactive with the rest of the visual system. The top down interaction, mediated by recurrent feedback connections, introduces contextual priors to influence the perceptual inference in the early visual cortex.

Early visual cortex

Neurons in the primary visual cortex are known to be tuned to specific elementary local features in the visual scenes. These features include location, line orientation, stereo disparity, movement direction , color and spatial frequency [1, 2]). It is also known that V1 neurons are also influenced by the surrounding context of the stimuli [3, 4, 5, 6]. The interpretations of the contextual modulations in these studies have been mostly limited to low-level mechanistic description in terms of facilitation and inhibition, or to subjective perceptual interpretations such as the neural correlate of pop-out or figure-ground saliency [4, 5, 7, 8]. Functionally, these modulations are thought to be related to computations of contours and saliency.

Some of the observed contextual modulations likely arise from the feedback mediated by the massive amount of recurrent connections from the extrastriate areas to V1. A plausible role of feedback is that of attentional selection based on the mechanisms of biased competition

[9]. The basic idea of biased competition is that when multiple stimuli are presented in a visual field, the different neuronal populations activated by these stimuli will engage in competitive interaction. Attending to a stimulus at a particular spatial location or to a particular object feature, however, could bias the competition in favor of the neurons representing the attended features or locations, enhancing their responses and suppressing the responses of the other neurons. However, the intra-cortical interaction in all biased competition models (e.g. [10]) is limited to lateral inhibition – a rather impoverished view on computation being done by the sophisticated machinery in the different visual areas.

Hierarchical Bayesian inference

Inspired by the recent experiments demonstrating top-down effect in V1, we have suggested that V1 can serve as a high-resolution buffer [7] that participates in many levels of visual computations through the recurrent feedback (see also Bullier's blackboard hypothesis [11]). In this context, a more appropriate theoretical framework for reasoning about top-down visual processing in the brain is that of Bayesian inference [12, 13, 14]. The idea can be traced back to the *unconscious inference* theory of perception by Helmholtz [15]. From the Bayesian perspective, the visual system arrives at the most probable interpretation of the visual scene by finding the *a posteriori* estimate S_i of the scene that maximizes $P(S_i|E,H)$, the conditional probability of a scene (S_i) given a particular sensory evidence (E), and the information you have already known (H), which, by Bayes' theorem, is given by,

$$P(S_i|E,H) = \frac{P(E|S_i,H)P(S_i|H)}{P(E|H)}$$

where $P(E|S_i,H)$ is the conditional probability of the evidence given the scene S_i and the prior information H. $P(S_i|H)$ is the prior probability of the scene given H, and $P(E|H)$ is the prior probability of the evidence given H. $P(E|S_i,H)$ can be factored into $P(E|S_i)P(H)$ if we assume H does not exert an direct effect on E.

This basic formulation can capture the interaction between two cortical areas, for example, V1 and V2. Let E be the evidence furnished to V1 by the retina (via LGN), processed with weights specified by $P(E|S_i)$, how well S_i can explain E. S_i is the output of V1 inference engine, and H a distribution of hypotheses generated by V2 based on its input from V1

as well as feedback from other higher order areas. V2 communicates to V1, but not LGN. Hence, the chain is Markovian. The feedback from V2 to V1 is given by the distribution of H weighted by $P(S_i|H)P(H)$, i.e. how well each hypothesis H can predict S_i. V1 is to find the S_i that maximizes $P(S_i|E,H) = P(E|S_i)P(S_i|H)P(H)/Z$. This scheme can then be applied again to V2 and V4 recursively and so on to generate the whole visual hierarchy. In this framework, each cortical area is an expert for inferring aspects of the visual scene, but its inference is constrained by both the data coming in and the top-down priors feeding back. Unless the image is simple and clear, each area normally cannot be completely sure of its inference, and has to harbor a number of hypotheses simultaneously. The feedforward input drives the generation of the hypotheses, the feedback from higher inference areas provides the priors to shape the inference at the earlier levels. Hierarchical Bayesian inference is concurrent across multiple areas, information does not flow forward to IT and then flow back to V1 and then back to IT. Such large loop would take too much time per iteration and is infeasible in real time inference. Rather, successive cortical areas in the visual hierarchy can constrain each other's inference in small loops instantaneously and continuously. The system, as a whole, might converge to an interpretation of the visual scene rapidly and simultaneously.

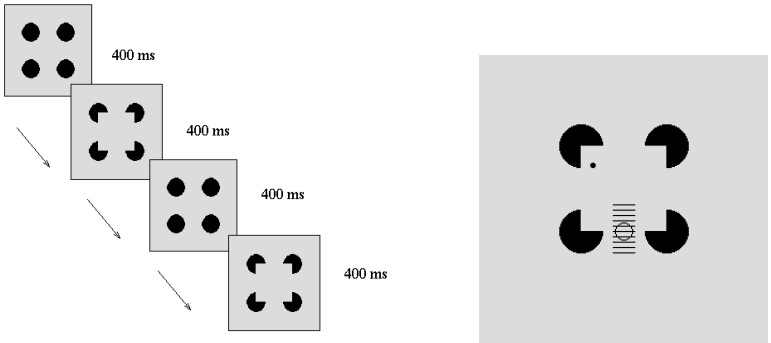
Evidence I: Subjective Contours

We carried out a series of neurophysiological experiments on awake behaving monkeys to test these ideas. The first experiment examined the neural representation of the famous Kanizsa illusion – a constructive inference created by the interaction of bottom-up surround contextual information and top-down priors. The second experiment examined a saliency effect that emerged from the brain's inference of 3D surface shape based on shading information. Both are wonderful case studies on the influence of higher order priors on early visual computation.

When viewing the display of stimulus sequence shown in Fig. 1a, we perceive a subjective square abruptly appear in front of four circular discs with vivid subjective borders even in regions of the image where there is no direct visual evidence for them. Could we see evidence of such illusion in the early visual cortex? Can we demonstrate that this illusion is generated by feedback from higher areas?

We [16] recorded the responses of over 200 V1 and V2 neurons of awake behaving monkeys to illusory figures in the course of one year. One or two neurons were tested per recording session. A number of stimuli (Fig. 1) were tested in each recording session. The most important stimuli is the illusory square (Fig. 1b), but many other stimuli, including the amodal figure (Fig. 1c), the stimuli with pac-men rotated (Fig. 1d) and a variety of real squares defined by contrast and lines (Fig. 1e and 1f), were also tested for control. Each stimulus was presented for 120 trials (10 conditions and 12 trials per condition). The monkey's task was to fixate a spot on the screen while stimuli were presented. In each trial, a sequence of four stimuli, 400 msec each, was presented. Figure 1a illustrates the presentation of the subjective square stimuli. First, four circular discs were presented. Then they were turned into pac-men, creating an illusion that a white square had abruptly appeared in front of the disks, occluding them. Over successive trials, the receptive field of the cell being recorded was placed at 10 different locations relative to the subjective contour or the corresponding positions in the other figures, 0.25^o apart, spanning a range of 2.25^o, as shown in Figure 1b. The receptive fields of the neurons, as plotted by a small oriented bar, was typically less than 0.8 degrees at that eccentricity (about 2 -3 degree away from the fovea). The gap between the pac-men was 2 degree wide. The neurons were considered to be sensitive to illusory contour if their response to the illusory contour, at the precise location of that contour, was significantly larger than their response to the amodal contour (Fig. 1c) and the conditions in which the pac-men were rotated (e.g. Fig. 1d).

a b



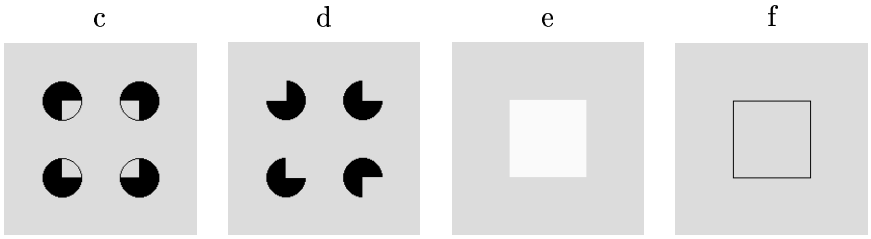


Figure 1: Selected stimuli in the subjective contour experiment. (a) An example sequence of stimulus presentation in a single trial. (b) Receptive field of the tested neuron was 'placed' at 10 different positions across the illusory contour, one per trial. (c) amodal contour – the subjective contour was interrupted by intersecting lines. (d) One of the several rotated pac-men controls. The surround stimulus was roughly the same, but there was no illusory contour. (e) One of the several types of real squares defined by luminance contrast. (f). Square defined by lines. These controls were used to assess the the neuron's positional sensitivity to real contour as well as for comparing the temporal responses between real and illusory contours.

We found that 26 percent of the V1 neurons in the superficial layer of V1 exhibited sensitivity to the illusory contour under our experimental paradigm. The neural correlate of the illusory contour signal emerged in a V1 neuron at precisely the same location where a line or luminance contrast elicited the maximum response from the cell (Fig. 2a). The response to the illusory contour was delayed relative to the response to the real contours by 55 ms (Fig. 2b), emerging about 100 ms after stimulus onset. The response to the illusory contour was significantly greater than the response to the controls, including the amodal contour or when the pac-men were rotated. At the population level, we found that sensitivity to illusory contours emerged at 65 ms in V2, and 100 ms in the superficial layer of V1 (Figures 2c and 2d). A possible interpretation of these data is that V2 detects the existence of an illusory contour by integrating information from a more global spatial context, and then generates a prior to facilitate the generation of contour inference in V1. Since the feedback connection is rather diffuse spatially, it likely only provides a general guidance in a spatially non-specific, but feature-specific manner, allowing the V1 circuitry to construct and complete a precise representation of the subjective contour.

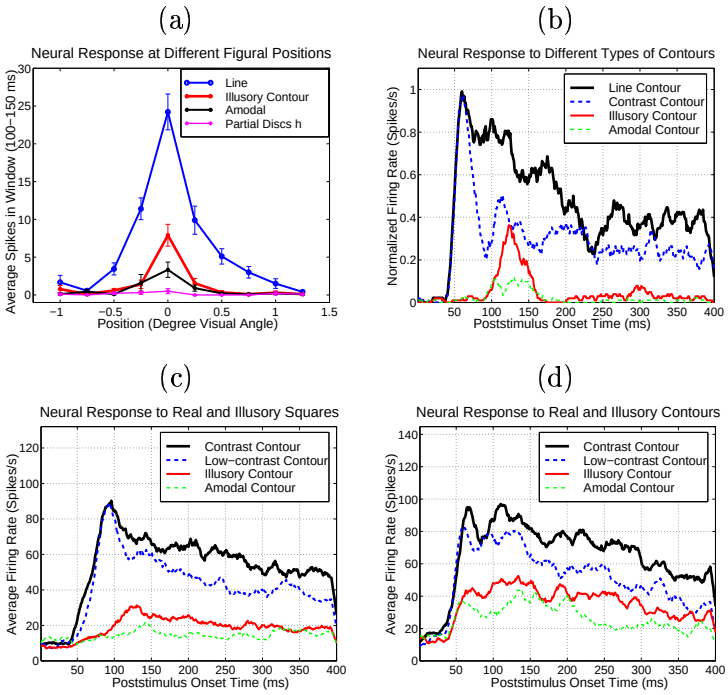


Figure 2: (a) The spatial profile of a V1 neuron's response to the contours of both real and illusory squares, in a temporal window 100-150 ms after stimulus onset. The real or illusory square was placed at different spatial locations relative to the receptive field of the cell. This cell responded to the illusory contour when it was at precisely the same location where a real contour evoked the maximal response from the neuron. This cell also responded significantly better to the illusory contour than to the amodal contour (T-test, $p < 0.003$) and did not respond much when the pac-men were rotated. (b) Temporal evolution of the cell's response to the illusory contour compared to its response to the real contours of a line square, or a white square, as well as to the amodal contour. The onset of the response to the real contours was at 45 ms, about 55 ms ahead the illusory contour response. (d) Population averaged temporal response of 50 V1 neurons in the superficial layer to the real and illusory contours.

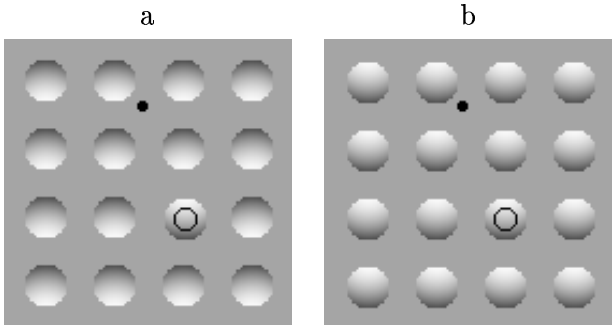
Evidence II: Shape from Shading

In the second experiment [8], we asked two questions. First, does 3D shape from shading information (a higher level information) can influence the processing in V1? Second, when we bias the monkey to look for a certain object, does this top-down bias have an impact on V1 inference?

We trained monkeys to perform a odd-ball detection task recorded from 550 V1 and V2 neurons while the monkeys were performing the fixation task. Shape from shading stimuli (Fig. 3a) are known to pop out, or readily segregate into different groups, while the 2D contrast patterns (WA, Fig 3e; WB, Fig. 3h) could not. The main difference between the two types of patterns is that the shading stimuli affords a 3D shape interpretation. Showing that V1 neurons are sensitive to shape from shading oddball and not the 2D contrast pattern oddball would establish V1 neurons are sensitive to 3D interpretation.

To evaluate whether the shape prior can influence the pop-out computation in the early visual cortex, we studied the response of V1 and V2 neurons to a variety of stimuli, in particular, the oddball condition and the uniform condition. In these two conditions of each type of stimuli, the receptive field of the tested neuron was covered by the identical stimulus element. An increase in neural responses to the odd-ball condition relative (e.g. Fig. 3a) to the uniform condition (e.g. Fig. 3b) can be considered a neural correlate of perceptual saliency.

For each stimulus type, four conditions (singleton, oddball, uniform and hole) were tested. The singleton stimulus and the hole stimulus were used as controls for each stimulus types. In the singleton stimulus, there was only one stimulus element, covering the receptive field. It was used to measure the neuronal response to direct stimulation of the RF alone, without any surround stimulus. The hole stimulus was the same as the uniform condition except the stimulus element on the receptive field was absent. It was used to measure the response to direct stimulation of only the extra-RF surround. In each trial, one of the conditions was displayed on the screen for 350 ms while the monkey fixated a red dot.



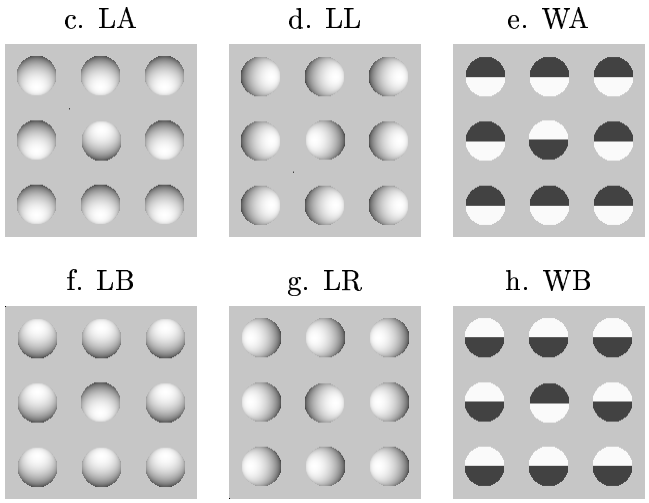


Figure 3: Higher order perceptual pop-out. We compared two conditions from each stimulus set: an oddball condition where the receptive field element is an oddball and a uniform condition where the RF element is one of the elements of the background. LA oddball (a) and LA uniform (b) conditions are shown for illustration. Six sets of stimuli were tested (c-h), i.e. lighting from above (LA), below (LB), left (LL) right (LR), and white above (WA) and white below (WB). In the actual experiment, a singleton stimulus (only the RF element) and a hole stimulus (background only, without the RF element) were also tested for each stimulus set for comparisons (see [8]).

We found that, after the monkeys were trained to detect the odd-ball in a stimulus, V1 and V2 neurons responded better to the odd-ball condition than the uniform condition for the LA or the LB stimuli (Figure 4a), but the difference in responses to the two conditions was weaker or absent in the WA or WB stimuli (Fig. 4b). These pop-out signals were found to be inversely correlated with the reaction time of the monkeys in detecting the oddball of the various types of stimuli (Fig. 4c,d), and hence could be considered a neural correlate of perceptual saliency of the oddball stimulus. Interestingly, before the odd-ball detection training, V2 but not V1 neurons exhibit sensitivity to shape from shading pop-out. This suggests that V2 may be the first cortical area where 3D inference is made about surface shape, and then the shape priors are fed back to V1 after the monkeys use the stimuli in their behavior. Another interesting observation is that when we change the relative presentation frequency of the stimuli to bias the monkeys' preference to a specific stimulus, for example, the LB oddball, the neural pop-out response becomes much stronger

for LB at the expense of LA. When we reversed the relative frequency, the pop-out response reversed correspondingly. Our interpretation is that when the monkey develops a preference of looking for a certain stimulus, the extrastriate ventral stream might provide a top-down feature (object) prior to facilitate the processing or detection of that particular stimulus in V1.

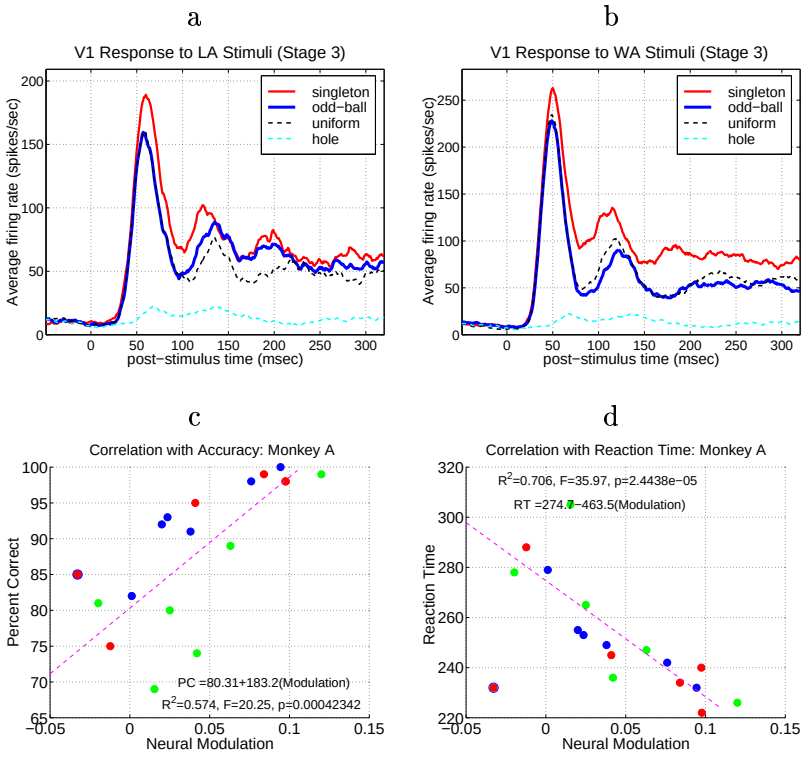


Figure 4: Temporal evolution of the normalized population average response of 30 V1 units from monkey A to the LA set (a) and the WA set (b) in a stage after the monkey had utilized the stimuli in its behavior. Each unit's response was first smoothed by a running average within a 15 ms window, then averaged across the population. A significant difference (pop-out response) was observed between the population average response to the oddball condition and that to the uniform condition in the LA set. No pop-out response was observed in the WA set. (c,d) The monkeys' behaviors and neural responses adapted after each stage of training. Here, behavior performance measurements (percent correct and reaction time) in three different training stages were regressed against the pop-out response. We found significant correlation between the neural pop-out responses and the behavioral performance (see [8] for details).

A new perspective

The findings of these two experiments support the the hierarchical Bayesian inference hypothesis of visual processing. Feedback from a higher order area to an earlier area can be conceptualized as providing top-down priors to bias the early inference. The impact of feedback is often subtle and becomes evident only when there is ambiguity in the visual stimuli, which is true in both of our experiments.

From this perspective, attention should not be conceptualized in terms of biased competition, but maybe more appropriately in terms of *biased inference*, or providing top-down priors in a hierarchical Bayesian inference framework. This conceptualization casts attention in a more mathematically tractable light. Feedback from the posterior parietal cortex could provide a spatial prior, i.e. prior expectation of how informative or interesting a particular visual location is. The influence of this spatial prior is called spatial attention. On the other hand, feedback from the ventral stream areas would provide a top-down object or feature prior, telling the early visual area what object the system is looking for, or what features we are expected to see in our high level inference of the existence of a particular object in the visual scene. This manifests as object attention or feature attention. The forms of attention depend on the task at hand, as different classes of priors would be required for different inferences. Theoretically speaking, priors can be derived from the statistics of stimuli and the processing constraints imposed by the computational tasks. Priors would form a bridge between behaviors, environments and perceptual inference. Understanding them should therefore be a central question in the study of biological vision.

References

[1] Hubel, D.H. & Wiesel,T.N. (1978). Functional architecture of macaque monkey visual cortex. *Proc. Royal Soc. B (London), 198*, 1-59.

[2] De Valois, R.L., & De Valois,K.K.(1988). *Spatial vision.* New York: Oxford University Press.

[3] Maffei, L., Fiorentini, A., The unresponsive regions of visual cortical receptive fields, Vision Research 16 (1976) 1131-1139.

[4] Knierim, J.J., Van Essen D.C., Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J. Neurophysiology 67 (1992) 961-980.

[5] Lamme, V.A.F., The neurophysiology of figure-ground segregation in primary visual cortex, J. Neuroscience 15(2) (1995) 1605-1615.

[6] Kapadia, M.K., Westheimer, G., Gilbert C.D., Spatial distribution of contextual interactions in primary visual cortex and in visual perception, J. Neurophysiol. 84(4) (2000) 2048-62.

[7] Lee, T.S., Mumford, D. Romero, R., Lamme, V.A.F., The role of the primary visual cortex in higher level vision. Vision Research 38(15-16) (1998) 2429-54.

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

[9] Desimone, R., Duncan, J., Neural mechanisms of selective visual attention, Annu. Rev. Neurosci. 18 (1995) 193-222.

[10] Deco, G., Lee, T.S., A unified model of spatial and object attention based on inter-cortical biased competition. Neurocomputing, (2002) (in press).

[11] Bullier, J. Integrated model of visual processing. Brain Res. Review, 36(2-3) (2001) 96-107.

[12] Mumford, D., On the computational architecture of the neocortex II, Biological Cybernetics 66 (1992) 241-251.

[13] Mumford, D., Pattery theory: a unifying perspective, in: Knill, D.C., Richards, W. (Eds), Perception as Bayesian inference, Cambridge University Press, 1996 pp. 25-62.

[14] Lee, T.S. (1995) A Bayesian framework for understanding texture segmentation in the primary visual cortex. *Vision Research, 35*, (18), 2643-2657.

[15] Helmholtz, H.V., Handbuch der physiologischen Optik, Leipzig: Voss, 1867.

[16] Lee, T.S. and Nguyen, M., Dynamics of subjective contour formation in the early visual cortex. Proc. Natl Acad. Sci. 98(4) (2001) 1907-1911.