Recurrent Processing in the Formation of Shape Percepts

Jan Drewes,^{1,2} Galina Goren,¹ Weina Zhu,^{3,4} and James H. Elder¹

¹Centre for Vision Research, York University, Toronto, Ontario, Canada M3J 1P3, ²Center for Mind/Brain Sciences, University of Trento, I-38068 Rovereto, Italy, ³School of Information Science, Yunnan University, Kunming 650091, China, and ⁴Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

The human visual system must extract reliable object information from cluttered visual scenes several times per second, and this temporal constraint has been taken as evidence that the underlying cortical processing must be strictly feedforward. Here we use a novel rapid reinforcement paradigm to probe the temporal dynamics of the neural circuit underlying rapid object shape perception and thus test this feedforward assumption. Our results show that two shape stimuli are optimally reinforcing when separated in time by ~60 ms, suggesting an underlying recurrent circuit with a time constant (feedforward + feedback) of 60 ms. A control experiment demonstrates that this is not an attentional cueing effect. Instead, it appears to reflect the time course of feedback processing underlying the rapid perceptual organization of shape.

Key words: recurrent processing; shape perception; temporal integration; feedback, grouping; perceptual organization; contour processing

Significance Statement

Human and nonhuman primates can spot an animal shape in complex natural scenes with striking speed, and this has been taken as evidence that the underlying cortical mechanisms are strictly feedforward. Using a novel paradigm to probe the dynamics of shape perception, we find that two shape stimuli are optimally reinforcing when separated in time by 60 ms, suggesting a fast but recurrent neural circuit. This work (1) introduces a novel method for probing the temporal dynamics of cortical circuits underlying perception, (2) provides direct evidence against the feedforward assumption for rapid shape perception, and (3) yields insight into the role of feedback connections in the object pathway.

Introduction

The human visual system is highly efficient at extracting salient information from complex visual scenes. For example, physiological and behavioral reactions to animals present in natural scenes occur within 150 ms of stimulus onset (Thorpe et al., 1996; Kirchner and Thorpe, 2006). Although it has been postulated that a feedforward sweep through the visual system is responsible for this very fast processing of visual information (Kirchner and Thorpe, 2006; Serre et al., 2007), it has also been suggested that recurrent or feedback processing may underlie certain aspects of rapid perception (Foxe and Simpson, 2002; Muggleton et al., 2011) and, in some cases, may even be required for visual stimuli

to reach conscious awareness (Pascual-Leone and Walsh, 2001; Fahrenfort et al., 2007).

Single-cell neurophysiology has provided clues to the connectivity, function, and dynamics of this recurrent processing in the monkey visual cortex. For example, Lamme (1995) found evidence for a late enhancement of response in primary visual cortex (V1) neurons, conditioned on the figure/ground status of the local stimulus. Because this information is signaled by texture boundaries well beyond the classical receptive field, this delayed modulation must arise either from recurrent interactions within V1 or through feedback connections from higher visual areas, in which receptive fields are larger. Zhou et al. (2000) found similar effects in V2 and V4, as well as V1. Using a reversible inactivation technique in anesthetized monkey, Hupé et al. (1998) were able to demonstrate that feedback from V5/middle temporal area (MT) serves to enhance the response of V1, V2, and V3 neurons to objects moving relative to background, particularly when the objects are less salient because of background clutter.

The fine temporal resolution of electroencephalography (EEG) methods has provided some insight into these recurrent processes in the human visual cortex. Bach and Meigen (1992) discovered a visual evoked potential negativity associated with

Received June 19, 2015; revised Oct. 21, 2015; accepted Nov. 16, 2015.

Author contributions: J.D. and J.H.E. designed research; J.D., G.G., and W.Z. performed research; J.D., W.Z., and J.H.E. contributed unpublished reagents/analytic tools; J.D., W.Z., and J.H.E. analyzed data; J.D. and J.H.E. wrote the paper.

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant and an NSERC CREATE training grant. W.Z. was supported by National Natural Science Foundation of China Grants 62263042 and 61005087 and a China Scholarship Council Grant.

The authors declare no competing financial interests.

Correspondence should be addressed to Jan Drewes, Center for Mind/Brain Sciences, University of Trento, 31 Corso Bettini, I-38068 Rovereto, Italy. E-mail: mail@jandrewes.de.

DOI:10.1523/JNEUROSCI.2347-15.2016

Copyright © 2016 the authors 0270-6474/16/360185-08\$15.00/0

the presence of texture boundaries, delayed with respect to the response to normal texture pattern onset, suggesting a later process responsible for segmentation (Caputo and Casco, 1999). Using more modern multichannel EEG methods and refined stimuli, Scholte et al. (2008) have been able to assess the dynamics of response to local boundary signals, as well as global contextual figure/ground information, separately localized to early visual areas versus higher areas in the parietal and temporal cortices. One of the most interesting findings is that contextual effects arise earlier in the temporal lobe than in the occipital lobe, suggesting feedback of this information from later to earlier visual areas of the object pathway.

More recently, transcranial magnetic stimulation (TMS) has allowed both the dynamics and causality of these recurrent circuits to be probed. Application of TMS over human MT+ (containing both area MT and the medial superior temporal area)/V5 is known to induce moving phosphenes in the visual field (Walsh and Cowey, 1998). Pascual-Leone and Walsh (2001) showed that subthreshold TMS over V1 can disrupt the perception of these phosphenes induced at V5. Importantly, this disruption peaks when the V1 TMS is applied after the V5 TMS, suggesting that feedback from V5 to V1 is causal on the phosphene percept.

More recently, TMS has been used to probe the recurrent circuits underlying figure/ground perception. Heinen et al. (2005) and Wokke et al. (2012) measured behavioral performance on a figure/ground task while varying the latency of TMS applied to the occipital lobe (V1 and possibly neighboring areas). They found two distinct latencies at which TMS was maximally disruptive, suggesting the sequencing of an early feedforward process, possibly coding boundary information, followed by an integration process reflecting surface segmentation. Camprodon et al. (2010) have used a similar method to probe the neural processes underlying the recognition of animals in natural scenes. They found that TMS applied over the occipital pole was maximally disruptive at two distinct latencies after stimulus onset, mostly consistent with the figure/ground results of Heinen et al. (2005).

Shape information is known to be critical for animal detection in natural scenes (Elder and Velisavljević, 2009). Although the classical receptive field of a V1 neuron can encode local shape information (e.g., orientation), it is blind to the figure/ground organization and global form of an object. The second phase of V1 sensitivity revealed by these experiments may in part reflect a feedback of this global shape information from higher visual areas in the object pathway, in which receptive fields are larger and neurons are tuned to more complex shape properties (Pasupathy and Connor, 1999; Kourtzi and Kanwisher, 2001). A TMS study by Wokke et al. (2013) addressed this hypothesis directly. In this study, observers discriminated between two illusory contour stimuli with slightly different shapes. On some trials, TMS was applied, either at the occipital pole to disrupt processing in V1/V2 or over the lateral occipital lobe to disrupt processing in the lateral occipital complex (LOC). Application of TMS was found to disrupt performance at both locations, but interestingly, disruption was maximal when TMS was applied later over V1/V2 than over LOC. This is strongly suggestive of a feedback process in the grouping of contour fragments to form shape percepts.

These studies rely on the disruptive effects of TMS on performance in a visual task to infer properties of the recurrent processes underlying object segmentation and shape perception. In this sense, the approach is cousin to classical backward masking methods (Enns and Di Lollo, 1997, 2000; Habak et al., 2006), in which visual noise is found to maximally disrupt shape process-

ing when applied after the target stimulus. One potential limitation of these methods is that the induced deficit in visual processing can potentially be attributable to a nonspecific factor. The transient onset of a strong visual mask or the repeated application of a focused magnetic field may generally disrupt processing and lower performance on a range of tasks by introducing noise or distracting attention, and thus an observed deficit does not demonstrate that the feedback process is specific to the visual task under study.

Here we posit that, if recurrent processing plays an important role in the rapid segmentation and perception of objects, it should also be possible to enhance performance by carefully sequencing visual stimuli that are informative for the task. This could lead to a new purely behavioral methodology with greater specificity to the task under study, in this case, the rapid perceptual organization of shape.

Materials and Methods

Experiment 1: rapid reinforcement

Natural scenes are brimming with diverse visual cues, including texture, color, and shape (Elder and Velisavljević, 2009; Crouzet and Serre, 2011). Each of these cues may take a different path through the visual system, involving separate recurrent circuits with unique time constants, and this diversity may obscure any behavioral expression of these recurrent processes. To avoid this, we restrict our attention here to the shape pathway, using natural stimuli that carry only shape information, embedded in a synthetic cluttered background. To probe the dynamics of the underlying shape processing circuit, we use a novel rapid reinforcement method in which shape stimuli are presented twice to the observer.

Figure 1 shows the logic behind this method. We posit a recurrent circuit in the human cortical object pathway in which global shape information, coded in LOC, for example, is fed back to earlier areas of visual cortex (e.g., V1) to guide contour extraction. [Although we use V1/LOC here as a concrete example, other visual areas, such as V2 and V4, could also be involved (see Discussion).] Let t_f denote the average time for shape information coded in V1 to feed forward to LOC, and let t_b denote the average time for information coded in LOC to feed back to V1. Consider a repetition paradigm in which a first presentation of the target stimulus is followed, after a delay, by a second, reinforcing presentation. If this interstimulus delay is of duration $t_f + t_b$, then the reinforcing information from the second stimulus presentation will arrive at V1 from the lateral geniculate nucleus (LGN) at the same time feedback information from the first stimulus presentation is arriving from the LOC. We hypothesize that this confluence of feedforward and feedback information may result in a peak in performance relative to shorter or longer delays, for which the feedforward and feedback of relevant stimulus information will not be synchronized at V1. By this logic, measuring shape discrimination performance as the delay between presentations is varied can yield insight into the nature of the neural circuits underlying rapid shape processing.

Stimuli. We used a shape discrimination paradigm in which observers must distinguish between natural animal shapes and “metamer” shapes that have been matched to the animal shapes in their local geometry.

Natural target shape stimuli were constructed from blue-screened images of 238 animal objects selected from the Hemera Photo-Object database. Each object boundary was represented as a 40-sided polygon. To simulate fragmentation caused by occlusions and contrast dropouts in natural scenes, every second side of the polygon was removed, resulting in a cycle of disconnected linear elements representing the shape (Fig. 2A, left column).

Metamer shapes were constructed from each animal contour by randomizing the sequence of turning angles between neighboring line segments, thus maintaining the local geometry of the shapes while removing all global shape properties (Fig. 2A, right column). As a result of this construction, to discriminate between a natural and a metamer stimulus, the observer must integrate multiple line elements into a coherent representation of shape.

Stimulus presentation was controlled using an Apple MacPro, running Mathworks MATLAB in conjunction with the Psychophysics Tool-

Figure 1. Simplified illustration of feedforward and feedback processing in the primate object pathway. TE, Temporal cortex; TEO, temporal-occipital area.

box (Brainard, 1997). Subjects were seated 60 cm from the display monitor (21-inch CRT, 100 Hz refresh rate) in a dimly lit room. Screen background luminance (black) measured <0.1 cd/m², whereas foreground luminance (white) measured 77.6 cd/m². At the 60 cm viewing distance, each line element was ~ 20 arcmin in length and 4 arcmin in width, and each shape subtended ~ 3 – 4° of visual angle.

The shape stimuli were embedded in a 1 s dynamic noise sequence consisting of randomly positioned and oriented line elements, matched in length to the line elements of the target shape, within a circular visual field of diameter 23.5° (Fig. 2D). Shape stimuli were presented at an eccentricity of 5.88° , at a random orientation and polar angle relative to fixation.

Subjects were instructed to maintain fixation on a central fixation cross for the duration of each trial.

Procedure. We ran two conditions: (1) a single presentation condition; and (2) a dual presentation condition. In the single presentation condition, the target shape was presented in only one 10 ms frame, after a random (uniformly distributed) delay of 250–500 ms after onset of the continuous dynamic noise. In the dual presentation condition, the target shape was presented a second time, at the same location and orientation, after a variable interstimulus interval (ISI) of 10–110 ms (sampled at 20 ms intervals). Note that at an ISI of 10 ms, the two 10 ms presentations appeared as a continuous 20 ms presentation, without a gap. The paradigm is shown in Figure 2, B and C.

We used a blocked design, in which the presentation condition (single/dual) varied between blocks and the ISI varied within dual-condition blocks. Before each block, subjects were informed whether to expect single or dual presentation. Within each presentation condition and ISI, half of the trials involved an animal target and half involved a metamer target, and these were interleaved randomly. The observer's task was simply to indicate with a button press which of the two types of target shape was present. Auditory feedback was given for errors.

For each presentation condition and ISI, 50 trials of an adaptive psychometric procedure (QUEST; Watson and Pelli, 1983) was used to estimate the number of background noise elements resulting in 75% correct performance.

One batch of 11 observers (aged 19–46 years, six males and five females, all reporting normal or corrected-to-normal vision) performed an average of 11 blocks each (range, 4–16) for each presentation condition and ISI.

After this experiment was completed, a follow-up control was conducted to assess any effect of eye movements. Four observers (aged 22–26 years, one male and three female, all reporting normal or corrected-to-normal vision) took part in a shorter form of the experiment (ISIs of 10, 50, and 110 ms, four blocks per observer) in which gaze location was monitored. Using an Eyelink 1000, saccades were detected during each trial. The built-in saccade detection mechanism was parameterized to the default thresholds of $9500^\circ/\text{s}^2$ acceleration and $30^\circ/\text{s}$ velocity.

Experiment 2: spatial cuing

To distinguish between the recurrent shape-processing hypothesis and a possible attentional cuing effect, we conducted a second dual presentation experiment in which the target shape in one of the two presentations is replaced by an attentional cue that provides no relevant shape information. There were four presentation conditions: (1) one in which an attentional cue was substituted for the first presentation of the target (Cue/Target); (2) one in which the cue was substituted for the second presentation of the target (Target/Cue); (3) one identical to the dual presentation condition of Experiment 1 (Target/Target); and (4) one identical to the single presentation condition of Experiment 1.

Three different cues were assessed for Cue/Target and Target/Cue conditions (see Fig. 4): (1) a cycle of 20 disconnected linear elements forming a circle shape, similar in size to the target shape. This cue has the advantage of being similar in its geometric properties to the target shape, thus serving as a good control. Conversely, it may not be as salient as a natural shape, and its similarity to the target shape could result in it serving as a mask rather than an attentional cue. For this reason we also assessed the following: (2) a large red X, matched in extent to the diameter of the average target stimulus (3°); and (3) a small red X, 0.6° in extent. Because of their color and the presence of an X junction, both of these cues should serve as highly salient attentional cues, and, given their distinctness from the target shape, we expect masking to be minimal, particularly for the smaller X.

We again used QUEST to estimate noise thresholds for shape discrimination, for three different ISIs (10, 50, and 110 ms).

The experiment consisted of three separate sessions, one for each cue, run in random order over subjects. Each session consisted of eight blocks, two for each presentation condition, run in random order. Before each block, observers were informed of the presentation condition they were about to run. One batch of five observers participated in Experiment 2 (aged 21–37 years, one male and four females), three of whom had also participated in Experiment 1.

Results

Experiment 1

Figure 3A shows average and individual results for our 11 observers. We assessed shape discrimination performance by measuring noise (number of oriented noise elements) thresholds at 75% correct performance. Although these thresholds are relatively low (on the order of 18 for the single presentation condition and 33 for the dual presentation condition), perceptual integration of the dynamic noise field over time resulted in a much higher perceived noise density.

Because there was considerable variation in the mean threshold for each observer, we have shifted each observer's threshold function vertically to match the mean of the grouped data, facilitating comparison of the shape of the curves. Dual presentation

Figure 2. Experiment 1 design. **A**, Example shape stimuli: animals (left column) and corresponding metamers (right column). **B**, Temporal sequence of the dual presentation condition. Boxes labeled N represent noise-only frames, and boxes labeled N + T represent frames with noise and target shape stimulus. The total duration of the stimulus was fixed at 1 s. **C**, Temporal sequence of the single presentation condition. **D**, Sample stimulus frame with animal target (N + T). TE, Temporal cortex; TEQ, temporal–occipital area.

Figure 3. Experiment 1 results. **A**, Performance on the shape discrimination task, in terms of threshold noise (number of distractor elements) at 75% correct. Red, Mean and SEM of the single presentation condition; blue, mean and SEM of the dual presentation condition; magenta, Gaussian fit. Pale gray lines represent individual subjects, shifted vertically to match the global average. **B**, Eye movement control experiment. Plots show percentage of trials in which observers made a saccade away from fixation, as a function of time relative to onset of the first target shape presentation. The vertical dashed line indicates the onset of the second target shape presentation.

of the target shape resulted in a clear improvement in discrimination performance (i.e., higher noise thresholds). To assess significance of this improvement for each observer, we averaged the dual presentation threshold estimates over ISI within each block. Separate independent-measures *t* tests for each observer of the resulting mean thresholds against the thresholds for the single presentation blocks confirmed a significant improvement in performance for all 11 observers ($p < .03$). To perform a group analysis, the thresholds for each observer computed above were averaged over blocks. A repeated-measures *t* test of these thresholds over observers confirmed the significance of the improvement for the group ($t_{(10)} = 11.25$, $p < 10e-6$).

However, the most interesting result is that, within the dual presentation condition, performance varied non-monotonically as a function of ISI, peaking at 50 ms delay between the two stimulus frames. To better quantify and assess the statistical significance of this result, we fit a raised Gaussian model to the data for threshold noise N (number of noise elements) as a function of ISI t :

$$N = a + b \exp\left(-\frac{(t - t_0)^2}{2\sigma^2}\right)$$

Fitting this model to the grouped data yielded a refined estimate of the ISI for peak performance of $t_0 = 61 \pm 8$ ms (68% confidence interval estimated by bootstrapping over observers).

A positive peak in threshold entails conditions on the derivatives of the model: the first derivative must be zero, and the second derivative must be negative for some positive ISI t_0 . For the raised Gaussian model, this requires that both b and t_0 are positive. Therefore, significance was assessed using a null hypothesis test on the joint values of these two parameters.

Specifically, the null hypothesis was that at least one of b and t_0 was zero or negative, and the alternate hypothesis was that both were positive. For the grouped data, the tests were performed by bootstrapping over observers. Using these criteria, the positive peak for the grouped data was found to be highly significant ($p < 0.0001$).

It is apparent from Figure 3A that most observers follow the trend of the grouped data, although there is variation attributable in part to the limited number of trials per observer. To assess this variation, we bootstrapped over blocks within each observer's data and fit the raised Gaussian model to each of 10,000 bootstrapped samples. Because we sampled ISI at 20 ms intervals, we constrained the dispersion parameter σ of the raised Gaussian model to be >10 ms to avoid degenerate fits. Based on the joint criteria on b and t_0 described above, all 11 observers showed a trend (i.e., more than half of bootstrapped trials) toward a positive peak in performance for intermediate ISI. For 6 of our 11 observers, this trend was significant at the 0.05 level, and for 9 of our 11 observers, it was significant at the 0.1 level. Thus, although there is considerable variation across observers, the peak in performance at intermediate ISI observed in the grouped data appears to be quite general and not driven by just one or two observers.

The variation in threshold as a function of ISI represents a 17% modulation relative to the continuous 20 ms presentation condition (i.e., 10 ms ISI). However, we stress that this does not represent the entire contribution of recurrent processing because recurrent processing is presumably operating normally in the single presentation condition as well, just not with the injection of a second stimulus. Given this, we do not think that the amplitude of the modulation should be taken as an index of the impor-

tance of the phenomenon. Rather, it is important that (1) there is a significant peak and (2) it occurs at ~ 60 ms ISI, which provides clues to the underlying recurrent circuit.

Are there alternative factors that could explain these results? One possibility is overt attention (eye movements). Observers were asked to maintain fixation during the entire stimulus interval, and the ISIs we used (<110 ms) are shorter than typical saccade latencies. However, under certain conditions, express saccades as short as 80–120 ms have been reported (Kingstone and Klein, 1993), and it is therefore conceivable that such saccades contributed somehow to our findings, e.g., by lowering performance for longer ISIs. To assess this, we conducted a shorter form of our experiment on four new observers (three ISIs, four blocks per observer) in which eye position was monitored. The proportion of trials in which saccades were made during the interval spanning the two target presentations was very low (between 0.4 and 2%), and no systematic variation with ISI was observed (Fig. 3B). Thus, overt attention does not appear to be responsible for our findings.

How about covert attention? Specifically, suppose that the first target presentation serves partially as a spatial cue, attracting attention to the relevant location in the visual field and thus priming the visual system for the second target presentation. Studies have estimated the time required for this attentional shift to be on the order of 50 ms (Bergen and Julesz, 1983; Koch and Ullman, 1985), roughly consistent with the ISI producing optimal shape discrimination performance in our first experiment.

To distinguish this attentional account from our recurrent shape-processing hypothesis, we require a control condition in which the first presentation of the target provides an attentional cue but does not provide relevant shape information. This brings us to our second experiment.

Experiment 2

Figure 4 shows the results grouped over observers. We found that, regardless of cue type, performance for the Cue/Target and Target/Cue conditions did not significantly exceed performance for the single presentation condition (Circle: $t_{(4)} = -1.12$, $p = 0.32$; X: $t_{(4)} = -0.99$, $p = 0.37$; x: $t_{(4)} = 1.04$, $p = 0.35$), and there was no sign that performance depended in any consistent way on the ISI. Conversely, for the Target/Target condition, we again found that performance was significantly better than for the single presentation condition ($t_{(4)} = 4.24$, $p = 0.0132$), and again there was a peak in performance near 50 ms ISI.

Although we were primarily concerned about spatial attention, attention is known to act also within feature dimensions (Treue and Martínez Trujillo, 1999). In our tasks, the targets comprise short line segments at various orientations. If featural attention were playing a role in our first experiment, we would expect to also see an increase in thresholds when cueing with the circle shape cue, which was formed from line segments matching the target in length, width, and separation, and spanning all of the orientations that comprise the target, but we saw no sign of this (Fig. 4). Evidently, the facilitation induced by the first presentation is more specific to shape.

These results thus rule out attention as the explanation for the rise in performance with the reinforcement paradigm and for the peak in performance near 60 ms ISI found in Experiment 1 but are consistent with the recurrent shape-processing hypothesis, namely that the peak in performance is attributable to the confluence of feedforward and feedback information arriving synchronously at an early stage in the cortical object pathway.

Discussion

Our experiments yield the surprising result that the extent of facilitation between two shape stimuli depends non-monotonically on the delay between their presentations, peaking at a delay of 60 ms. This is strong evidence for a recurrent circuit underlying shape processing in the human cortical object pathway and is, to our knowledge, the first time such an effect has been found. However, these results are consistent with a number of previous backward masking studies showing that performance can be hindered by the presentation of a masking stimulus after presentation of the target. Using simple geometric shapes and either metacontrast or four-dot masking, Enns and Di Lollo (1997, 2000) found such backward masking to peak in effectiveness when the mask was presented 45–90 ms after the target. More recently, Habak et al. (2006) have found backward masking for shape discrimination to peak 53–107 ms after stimulus onset. This is consistent with the recurrent shape-processing hypothesis: if irrelevant or contradictory information arrives at V1 from the LGN just as relevant feedback cues are arriving from higher visual areas such as LOC, we would expect a performance decrement.

Recent TMS studies also provide evidence for feedback in the human object pathway. Applied to early visual areas, TMS blocks the perception of briefly presented stimuli when applied 30 ms before stimulus onset and up to 50 ms after stimulus onset (Corthout et al., 1999). Intriguingly, there appears to be a second, later time window, 80–120 ms after stimulus onset (Walsh and Cowey, 1998; Lamme and Roelfsema, 2000) at which TMS is effective, again suggesting a role for feedback, with a round-trip time delay of 30–150 ms.

Numerous studies have suggested an involvement of feedback from temporal areas to V1 and V2 in the formation of illusory shape percepts (Murray et al., 2001; Halgren et al., 2003; Yoshino et al., 2006), but a more recent TMS study (Wokke et al., 2013) provides perhaps the most direct evidence for the causal role of feedback in the perception of shape. In this study, observers were to discriminate between two illusory contour stimuli with slightly different shapes. On some trials, TMS was applied, either at the occipital pole to disrupt processing in V1/V2 or in the lateral occipital lobe to disrupt processing in the LOC. Application of TMS was found to disrupt performance at both locations, but interestingly, the effect depended critically on the timing. In LOC, TMS disrupted processing when the pulse occurred 100–122 ms after stimulus onset, whereas in V1/V2, processing was disrupted when the pulse was applied later, 160–182 ms after stimulus onset. This is strongly suggestive of a feedback process in the grouping of

Figure 4. Experiment 2, spatial cueing control. In two variations of the dual presentation condition of Experiment 1, one of the two shape targets was replaced with a spatial cue, in either Cue/Target (left column) or Target/Cue (middle column) order. Row 1, Circle cue; row 2, large X cue; row 3, small x cue; row 4, average over cue types; red, mean and SEM of the single presentation condition; blue, mean and SEM of the dual presentation condition. The panel on the bottom right replicates the dual presentation (Target/Target) condition of Experiment 1.

contour fragments to form shape percepts, with a one-way feedback time constant (LOC to V1/V2) of 40–80 ms.

Given that this delay should reflect only the feedback stage of processing, this time constant of 40–80 ms is a little long relative to the 60 ms time constant estimated from our reinforcement paradigm or the time constants ranging from 45 to 107 ms estimated from previous backward masking paradigms, both of which should reflect “round-trip” (feedforward + feedback) processing. It is possible that the relatively large gaps between illusory shape inducers (3.23°) led to longer computation times. Parametric spatial manipulation of the stimuli to vary eccentricity and inducer separation could serve to test this hypothesis.

Both the earlier backward masking studies and more recent TMS studies probe recurrent cortical circuits by inducing a deficit in stimulus processing. Conversely, our reinforcement paradigm demonstrates an enhancement in stimulus processing. This is an important distinction, because an induced deficit in visual processing can potentially be attributable to a nonspecific factor. The transient onset of a strong visual mask or the repeated application of a focused magnetic field may generally disrupt processing and lower performance on a range of tasks by introducing noise or distracting attention, and thus an observed deficit does not demonstrate that the feedback process is specific to the visual task under study (shape processing, in our case). Using our reinforcement paradigm, we show that enhancement does not occur for arbitrary spatial cues but only when the second presentation carries detailed shape information relevant to the task. This suggests that the recurrent process underlying our results is not a general attention or noise attenuation mechanism but a recurrent circuit specific to the perceptual organization of shape.

Based in part on previous work (Murray et al., 2001; Halgren et al., 2003; Yoshino et al., 2006; Wokke et al., 2013), we have suggested that this recurrent circuit could entail feedback from the LOC to the early visual cortex, but of course our behavioral method cannot by itself confirm this: other forms of corticocortical feedback could also be playing a role.

The recurrent circuit connecting areas V1 and V2 may be one of the other circuits involved, but this circuit is unlikely to account for all of our results. First, our shape stimuli were 3–4° in extent and presented at an eccentricity of 5.88°. At this eccentricity, human V2 receptive fields are estimated to average ~2° in diameter (Dumoulin and Wandell, 2008), too small to code for the global shape of our targets. On the other hand, although the metamer shapes match the animal shapes perfectly in the pairwise geometry of neighboring oriented elements, all higher-order geometries, starting from local triplets of oriented elements, are not matched. Thus, completely global integration of the elements comprising the shape stimuli may not be required to perform the task, because even a local triplet of oriented elements provides a statistical cue for discriminating animal from metamer shapes.

However, a second issue concerns the timing. The median latency of visual response is estimated to be ~60 ms in V1 and 85 ms in V2 (Bullier, 2001b). Because V1 and V2 are very close together, conduction times are short, estimated to be on the order 1–2 ms (Girard et al., 2001). This would predict that the latency of feedback would be closer to 30 ms than to the 60 ms we observe. Although mean threshold is slightly higher for 30 ms ISI relative to 10 ms ISI, this only holds for 6 of our 11 observers (Fig. 3A) and is not significant for the group ($t_{(10)} = 0.79$, $p = 0.45$).

In addition to LOC, other visual areas, such as V4 and MT/V5, have neurons with receptive fields large enough to span our shape stimuli. Indeed, MT/V5 has been shown to play a consequential role in feedback to V1 in a number of studies (Pascual-Leone and Walsh, 2001). However, median latency of visual response for MT neurons has been estimated at only 75 ms (Bullier, 2001b), and this would predict feedback effects that peak for much smaller ISIs. Consistent with this prediction, using a dual-pulse TMS paradigm, Pascual-Leone and Walsh found that MT to V1 feedback effects peaked at 25 ms delay. For these reasons, we believe that higher visual areas in the object pathway (e.g., V4 and LOC) are most likely to account for the 60 ms peak we observe in our data.

In addition to corticocortical feedback, could recurrent circuits within visual areas be playing a role? A comparison between the facilitation time constant of 60 ms estimated with our dual presentation method against known physiological dynamics can

provide some guidance here. It is known that the orientation tuning of layer 4B cells of V1 changes dramatically over time, and this may arise in part from horizontal recurrent interactions within V1 (Ringach et al., 1997). However, the delay between peak excitation and inhibition is on the order of 20 ms, too short to match our behavioral findings.

It is known that there are also long-range interactions within V1. Our stimuli subtended 3–4° of visual angle and were presented at 5.88° eccentricity, where cortical magnification is ~0.4 mm/° (Duncan and Boynton, 2003). Given cortical horizontal propagation speeds of ~0.14 m/s (Bringuier et al., 1999), this predicts a delay of ~10 ms for information to travel from one end of the stimulus to the other. Given that multiple iterations of long-range information exchange may be required before the network settles, these calculations cannot rule out a contribution of long-range interactions within early visual areas.

Distinguishing this from corticocortical feedback from a higher visual area using purely behavioral methods is a challenge, but recent physiological work suggests one possible avenue. Sugihara et al. (2011) measured border-ownership signals in macaque V1 and V2 as a function of the distance of the receptive field from the contextual component of the stimulus that distinguished figure/ground polarity. Although they found that the border-ownership signals were delayed at larger distances, the delay was far too small to be accounted for by horizontal signal propagation. From these results, they inferred that border-ownership signals were more likely to be the result of feedback from higher visual areas.

An analogous behavioral method is conceivable. In particular, it may be possible to devise a version of our dual presentation paradigm in which the retinal distance that must be spanned to compute the relevant shape information can be varied. As for the physiological border-ownership experiments, smaller effects of retinal distance would favor the feedback account.

Notes

Supplemental material for this article, featuring an animation of the dynamic visual stimulus used in the experiments reported here, is available at <http://elderlab.yorku.ca/recurrent2015>. This material has not been peer reviewed.

References

- Bach M, Meigen T (1992) Electrophysiological correlates of texture segregation in the human visual evoked potential. *Vision Res* 32:417–424. [CrossRef Medline](#)
- Bergen JR, Julesz B (1983) Parallel versus serial processing in rapid pattern discrimination. *Nature* 303:696–698. [CrossRef Medline](#)
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436. [CrossRef Medline](#)
- Bringuier V, Chavane F, Glaeser L, Frégnac Y (1999) Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* 283:695–699. [CrossRef Medline](#)
- Bullier J (2001b) Integrated model of visual processing. *Brain Res Rev* 36:96–107. [CrossRef Medline](#)
- Camprodon JA, Zohary E, Brodbeck V, Pascual-Leone A (2010) Two phases of V1 activity for visual recognition of natural images. *J Cogn Neurosci* 22:1262–1269. [CrossRef Medline](#)
- Caputo G, Casco C (1999) A visual evoked potential correlate of global figure-ground segmentation. *Vision Res* 39:1597–1610. [CrossRef Medline](#)
- Corthout E, Uttl B, Walsh V, Hallett M, Cowey A (1999) Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *Neuroreport* 10:2631–2634. [CrossRef Medline](#)
- Crouzet SM, Serre T (2011) What are the visual features underlying rapid object recognition? *Front Psychol* 2:326. [CrossRef Medline](#)
- Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. *Neuroimage* 39:647–660. [CrossRef Medline](#)

- Duncan RO, Boynton GM (2003) Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron* 38:659–671. [CrossRef Medline](#)
- Elder JH, Velisavljević L (2009) Cue dynamics underlying rapid detection of animals in natural scenes. *J Vis* 9:7. [CrossRef Medline](#)
- Enns JT, Di Lollo V (1997) Object substitution: a new form of masking in unattended visual locations. *Psychol Sci* 8:135–139. [CrossRef](#)
- Enns JT, Di Lollo V (2000) What's new in visual masking? *Trends Cogn Sci* 4:345–352. [CrossRef Medline](#)
- Fahrenfort JJ, Scholte HS, Lamme VA (2007) Masking disrupts reentrant processing in human visual cortex. *J Cogn Neurosci* 19:1488–1497. [CrossRef Medline](#)
- Foxe JJ, Simpson GV (2002) Flow of activation from V1 to frontal cortex in humans. *Exp Brain Res* 142:139–150. [CrossRef Medline](#)
- Girard P, Hupé JM, Bullier J (2001) Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J Neurophysiol* 85:1328–1331. [Medline](#)
- Habak C, Wilkinson F, Wilson HR (2006) Dynamics of shape interaction in human vision. *Vision Res* 46:4305–4320. [CrossRef Medline](#)
- Halgren E, Mendola J, Chong CD, Dale AM (2003) Cortical activation to illusory shapes as measured with magnetoencephalography. *Neuroimage* 18:1001–1009. [CrossRef Medline](#)
- Heinen K, Jolij J, Lamme VA (2005) Figure-ground segregation requires two distinct periods of activity in V1: a transcranial magnetic stimulation study. *Neuroreport* 16:1483–1487. [CrossRef Medline](#)
- Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394:784–787. [CrossRef Medline](#)
- Kirchner H, Thorpe SJ (2006) Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res* 46:1762–1776. [CrossRef Medline](#)
- Kingstone A, Klein R (1993) What are human express saccades? *Percept Psychophys* 54:260–273. [Medline](#)
- Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol* 4:219–227. [Medline](#)
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506–1509. [CrossRef Medline](#)
- Lamme VA (1995) The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci* 15:1605–1615. [Medline](#)
- Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 23:571–579. [CrossRef Medline](#)
- Muggleton NG, Banissy MJ, Walsh VZ (2011) Cognitive neuroscience: feedback for natural visual stimuli. *Curr Biol* 21:R282–R283. [CrossRef Medline](#)
- Murray RF, Sekuler AB, Bennett PJ (2001) Time course of amodal completion revealed by a shape discrimination task. *Psychon Bull Rev* 8:713–720. [CrossRef Medline](#)
- Pascual-Leone A, Walsh V (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292:510–512. [CrossRef Medline](#)
- Pasupathy A, Connor CE (1999) Responses to contour features in macaque area V4. *J Neurophysiol* 82:2490–2502. [Medline](#)
- Ringach DL, Hawken MJ, Shapley R (1997) Dynamics of orientation tuning in macaque primary visual cortex. *Nature* 387:281–284. [CrossRef Medline](#)
- Schulte HS, Jolij J, Fahrenfort JJ, Lamme VA (2008) Feedforward and recurrent processing in scene segmentation: electroencephalography and functional magnetic resonance imaging. *J Cogn Neurosci* 20:2097–2109. [CrossRef Medline](#)
- Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci U S A* 104:6424–6429. [CrossRef Medline](#)
- Sugihara T, Qiu FT, von der Heydt R (2011) The speed of context integration in the visual cortex. *J Neurophysiol* 106:374–385. [CrossRef Medline](#)
- Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522. [CrossRef Medline](#)
- Treue S, MartínezTrujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–579. [CrossRef Medline](#)
- Walsh V, Cowey A (1998) Magnetic stimulation studies of visual cognition. *Trends Cogn Sci* 2:103–110. [CrossRef Medline](#)
- Watson AB, Pelli DG (1983) QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys* 33:113–120. [CrossRef Medline](#)
- Wokke ME, Sligte IG, Steven Scholte H, Lamme VA (2012) Two critical periods in early visual cortex during figure-ground segregation. *Brain Behav* 2:763–777. [CrossRef Medline](#)
- Wokke ME, Vandenbroucke AR, Scholte HS, Lamme VA (2013) Confuse your illusion feedback to early visual cortex contributes to perceptual completion. *Psychol Sci* 24:63–71. [CrossRef Medline](#)
- Yoshino A, Kawamoto M, Yoshida T, Kobayashi N, Shigemura J, Takahashi Y, Nomura S (2006) Activation time course of responses to illusory contours and salient region: a high-density electrical mapping comparison. *Brain Res* 1071:137–144. [CrossRef Medline](#)
- Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. *J Neurosci* 20:6594–6611. [Medline](#)