

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

The Neural Representation of Multiple Objects in the Primate Visual System

 Danique Jeurissen,  Anne F. van Ham, and  Matthew W. Self

Department of Vision and Cognition, Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, 1105 BA Amsterdam, The Netherlands

Review of Gilad and Slovin and Martin and von der Heydt

One of the main tasks of the visual system is to combine the edges and surfaces of individual objects into a perceptual group, and thus create a representation of visual scenes in which multiple objects are segregated from the background. Many studies have focused on how single objects (usually simple figures, like small squares) are segregated. These studies have shown that the firing rate of a given cell in early visual cortex is greater when its receptive field (RF) overlaps with a figure than when it falls on a background. The increase in firing rate is referred to as figure–ground modulation (Lamme, 1995). Cells in early visual areas also encode the borders of figures. These cells often show a preference for a particular border of a figure (e.g., the left or right side) in their RF. This preference is termed border-ownership selectivity (Qiu et al., 2007). Recent research has begun to investigate how more realistic natural scenes with multiple, complex objects are represented in the primate visual system, how figure–ground modulation and border-ownership selectivity interact with each other, and how attention influences these signals. So far,

two main hypotheses have been proposed to explain the neural representation of multiple objects in the primate visual system. Early visual areas could either use a “rate code,” in which different figures are encoded by different firing rates, or a “synchrony code,” in which neurons that represent the same object fire in synchrony. Two exciting recent studies have examined how multiple objects in the visual scene are segregated from their background. Gilad and Slovin (2015) suggest that separate objects are labeled by different levels of activity in V1. Martin and von der Heydt (2015) suggest that feedback from cells in higher visual areas influences the synchrony between cells in lower area: cells which receive feedback from a common source are more likely to fire synchronous spikes. The authors test this idea by measuring the synchrony between cells representing the borders of objects to study their connectivity. Both studies make use of recordings in the visual cortex of awake, behaving monkeys performing attention-demanding behavioral tasks.

Gilad and Slovin (2015) measured the population response of V1 neurons with voltage-sensitive dye imaging (VSDI) to investigate whether a response amplitude code might be used to label multiple objects in the visual scene. Such a response amplitude code may be due to different firing rates for neurons that code for different objects. They presented monkeys with two motion-defined horizontal bars, and on some trials, the bars were con-

nected with semicircular segments to form a single object. The monkeys were trained to report whether the bars were separated or connected (Gilad and Slovin, 2015, their Fig. 1). The bars were always presented at the same locations, and VSDI measurements of the cortical surface in V1 were made at the retinotopic locations corresponding to the centers of the top and bottom bars. Importantly, the connecting segments fell well outside the imaged region and the classic receptive fields of the recorded neurons. Surprisingly, responses of neurons representing the top bar were consistently higher and responses of neurons representing the bottom bar were consistently lower when the bars were separated than when they were connected. This phenomenon is referred to as figure–figure modulation (ΔFF , i.e., the difference in the mean population response between the top and bottom bar when the bars are separated vs connected). This modulation was found across many experimental manipulations, such as varying figure sizes, having a connector at one or two sides of the figure, and motion directions for the connectors that vary in their motion-direction agreement with the bars. In a subset of experiments, the saliency of the connected percept was varied by decreasing the difference in motion direction between the connected segments and the background. The results show an excellent correlation between the monkeys' ability to perform the task and ΔFF : the monkeys' accuracy

Received July 2, 2015; revised Aug. 12, 2015; accepted Aug. 18, 2015.

We thank Pieter Roelfsema and the other members of the department of Vision and Cognition at the Netherlands Institute for Neuroscience for discussions and helpful feedback.

The authors declare no competing financial interests.

Correspondence should be addressed to Danique Jeurissen, Netherlands Institute for Neuroscience, Department of Vision and Cognition, Meibergdreef 47, 1105 BA Amsterdam, The Netherlands. E-mail: d.jeurissen@nin.knaw.nl.

DOI:10.1523/JNEUROSCI.2527-15.2015

Copyright © 2015 the authors 0270-6474/15/3512612-03\$15.00/0

to determine whether the bars were connected was higher when ΔFF (the difference in activity between the top and bottom bar) was larger (Gilad and Slovin, 2015, their Fig. 6). Indeed, ΔFF could be used to predict the choice of the monkey on single trials. The results suggest a very robust modulation (ΔFF) between the cortical representations of top and bottom bars when they form separate objects.

Gilad and Slovin (2015) interpret ΔFF to be a possible response amplitude code that could be used to label different objects in the visual scene. This exciting idea raises the questions of how the amplitude label is assigned to an object and how many unique amplitudes can be used to code multiple objects. In these experiments, the top bar was always presented closer to fixation than the bottom bar. The amplitude label might therefore be assigned based on the average eccentricity of the object, but this would pose a problem if multiple objects are placed at similar eccentricities in the visual field. Alternatively, the label might depend on how the animal directs its attention to the stimulus to solve the task. One interesting possibility discussed by Gilad and Slovin (2015) is that the monkeys solved the task by preferentially directing their attention to the top bar. Such a strategy could explain the consistent increase in response to the top bar relative to the bottom when the bars were disconnected. When the bars were connected, the modulation might spread from the top bar throughout the entire object, leading to an intermediate response amplitude over the cortical surface. Because an attentional signal would need time to spread from the top bar to the bottom bar, with longer bars requiring more time (Pooremaeli and Roelfsema, 2014), a latency analysis could reveal whether modulation based on attention is plausible. Future research in which animals direct their attention toward or away from the figures would help to determine if the modulation observed is a label for different figures or whether it reflects attentional selection of one figure and suppression of all others.

Gilad and Slovin (2015) also investigated whether synchrony could be used as a label to differentiate between the two stimuli. Although synchrony, calculated as the correlation coefficient between the VSDI signals on single trials after subtracting the mean signal, significantly increased when the bars were connected compared with when they formed two separate objects, the synchrony difference was very small and was only present in a narrow time window (120–160 ms). Moreover, it was less consistent over recording sessions and could not discriminate be-

tween conditions at the single-trial level. Gilad and Slovin (2015) therefore suggest that the amplitude difference is the best neural code for discriminating between separated and connected figures and the synchrony difference may hold additional, less consistent, figure information. However, the lack of finding a difference in synchrony between the conditions may be due to the recording techniques (VSDI vs spiking activity), because the synchrony may be present in high-frequency bands but not in the low-frequency bands that dominate the VSDI signal.

The role of spike synchrony in labeling objects was addressed in the study by Martin and von der Heydt (2015) in the context of an experiment that tested synchrony between pairs of neurons that code for the edges of objects. They examined the hypothesis that grouping of edges into a single coherent object is accomplished by feedback from “grouping cells” in higher visual areas to cells in lower visual areas that encode the borders of the object. A key prediction of this theory is that spatially separated cells in V1 and V2 that encode parts of the same object and have consistent border-ownership preferences (i.e., border-ownership selectivity that points toward the interior of the object) should receive a common input from the same grouping cell (Martin and von der Heydt, 2015, their Fig. 1). This common input would increase the likelihood of the cells firing within a brief time window, leading to an increase in synchrony. Martin and von der Heydt (2015) therefore used synchrony as a diagnostic tool to test for common input. They also more generally tested whether synchrony is enhanced when cells respond to the same object compared with when they respond to different objects, as would be predicted by theories of binding-by-synchrony (von der Malsburg, 1981; Singer and Gray, 1995). In this technically challenging experiment, they simultaneously recorded pairs of V1 and V2 neurons that had RFs on borders of the same figure (bound) or on two different figures (unbound) while the monkey directed its attention to the figure in the RF (attend) or away from the RF (ignore) (Martin and von der Heydt, 2015, their Fig. 2).

Martin and von der Heydt (2015) first identified the border-ownership selectivity of the two cells and classified this as either consistent, when they point toward each other (same grouping circuit cells), or inconsistent, when they pointed away from each other (different grouping circuit cells). They found that the spike synchrony (defined as the ex-

cess number of coincident spikes compared with that expected by chance in a time-window of 40 ms) between pairs of neurons with consistent border-ownership selectivity was higher when the RFs fell on the same object (bound-ignore condition) than when they fell on two different objects (unbound-ignore condition). The increased synchrony in the same grouping circuit is in line with the hypothesis that grouping cells in higher visual areas enhance activity of cells with an RF on a figure in lower visual areas. Thus, synchrony reflects connectivity between cells in the same grouping circuit, and it is this connectivity that eventually leads to the enhancement of feature responses with binding and attention. The authors found only weak support for the binding-by-synchrony hypothesis. Synchrony was weakly increased when the cells responded to the same rather than different objects, but the increase was only 0.6 Hz on a background coincidence rate of ~ 40 Hz, which is unlikely to provide a meaningful code. One caveat that should be mentioned is that measures of synchrony are sensitive to differences in the mean firing rate of the cells, because the number of coincidences will increase as the firing rate increases. Given that the mean firing rate was higher in the bound than in the unbound condition, this could potentially lead to an artificial increase in measured synchrony in this condition. The authors addressed this concern by subtracting the mean firing rate of each condition in their calculation of synchrony, which provides a measure of protection against this problem, as long as the firing rate differences between conditions are not too extreme.

The grouping-cell theory predicts that synchrony resulting from common input should occur at near-zero time-lags, because feedback projections have fast conduction velocities and should provide nearly simultaneous input to spatially separated cells. The peaks in the covariograms were relatively broad (~ 40 ms), making it difficult to determine the precise lag of the synchrony. To test whether grouping cells might induce synchrony that is more precise than this 40 ms time window, the authors performed an analysis in which they shuffled the timing of the spikes within short time windows and recomputed synchrony. If the measured synchrony relied on the precision of spike-timing at timescales shorter than

the window, it would be destroyed by shuffling the spikes, but if synchrony was only present at longer timescales, it would remain unaffected. The results showed that the precise timing of the spikes was critical for the measured synchrony when the spikes were shuffled within a 20 ms window, which indicates that synchrony was present at very brief timescales. Furthermore, the level of tight synchrony did not depend on the spatial separation or orientation tuning of the two cells, implying a feedback-based mechanism.

Martin and von der Heydt (2015) also addressed the effect of attention on synchrony. The authors had previously proposed that grouping cells might provide an efficient target for attentional selection. By enhancing the activity of a grouping cell, attention could select all the features that belong to the same object through the feedback connections made by these cells. One might then hypothesize that attention would enhance synchrony between cells that receive the same grouping cell input. This was not what the authors found, however. For pairs with consistent border-ownership tunings, synchrony was decreased when the figure was attended (bound-ignore vs bound-attend), whereas

for inconsistent pairs, synchrony was increased when the figure was attended. While this result remains puzzling given previous studies showing increases in spike–spike synchrony (Fries et al., 2008) and spike–field coherence with attention (Fries et al., 2001), the authors provide a possible interpretation: if attention acts via grouping cells it may enhance the activity of cells encoding the attended object and suppress nonattended distractors. The increased synchrony in the bound-ignore condition would then be a result of the common suppression received by cells encoding the nonattended object. This interesting proposal warrants further research.

In summary, these two papers use challenging techniques to advance our understanding of how the visual system groups together parts of the same object and labels different objects in the scene. Both papers suggest that objects are labeled through enhanced firing, and that synchrony does not provide a reliable label. They raise interesting questions about the interaction between attention and grouping that should be pursued by future research and bring us closer to understanding the mechanisms by which the binding problem is solved by the primate visual system.

References

- Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291:1560–1563. [CrossRef Medline](#)
- Fries P, Womelsdorf T, Oostenveld R, Desimone R (2008) The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4. *J Neurosci* 28:4823–4835. [CrossRef Medline](#)
- Gilad A, Slovin H (2015) Population responses in V1 encode different figures by response amplitude. *J Neurosci* 35:6335–6349. [CrossRef Medline](#)
- Lamme VA (1995) The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci* 15:1605–1615. [Medline](#)
- Martin AB, von der Heydt R (2015) Spike synchrony reveals emergence of proto-objects in visual cortex. *J Neurosci* 35:6860–6870. [CrossRef Medline](#)
- Pooresmaeili A, Roelfsema PR (2014) A growth-cone model for the spread of object-based attention during contour grouping. *Curr Biol* 24:2869–2877. [CrossRef Medline](#)
- Qiu FT, Sugihara T, von der Heydt R (2007) Figure-ground mechanisms provide structure for selective attention. *Nat Neurosci* 10:1492–1499. [CrossRef Medline](#)
- Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18:555–586. [CrossRef Medline](#)
- von der Malsburg C (1981) The correlation theory of brain function. In: *MPI biophysical chemistry, internal report 81-2*. Reprinted in *Models of neural networks II* (1994) (Domany E, van Hemmen JL, Schulten K, eds.). Berlin: Springer.