

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.

When Feature-Based Attention Distorts Neural Representations

Christian Kaul^{1,2} and Alex L. White¹

¹Department of Psychology and ²Center for Neural Science, New York University, New York, New York 10003
Review of Zirnsak and Hamker

Imagine you are watching a fox running through a dense forest. To keep track of him as he disappears and reappears, you may focus on all things that are reddish and moving to the right. Feature-based attention selectively enhances the processing of a particular visual feature value within a dimension, such as a particular color or motion direction. In the context of feature-based attention to motion directions specifically, the underlying neural mechanisms have been explored with electrophysiological recordings from the middle temporal area in monkeys (Treue and Martínez Trujillo, 1999; Martínez-Trujillo and Treue, 2004). These previous studies led to the formulation of the feature-similarity gain model of up-modulations of activity in neurons that prefer the attended feature value and down-modulations of activity in neurons that do not prefer the attended feature value. These modulations enhance the population response to stimuli with the attended direction. Interestingly, the effect spreads involuntarily across space: feature-based modulations occur for cells with receptive fields at distant, task-irrelevant locations.

Several studies have measured the effect of feature-based attention with psychophysics (Alais and Blake, 1999; Saenz et al., 2003;

Boynton et al., 2006; Ling et al., 2009), blood oxygen-level dependent (BOLD) signals (Saenz et al., 2002; Serences and Boynton, 2007), and event-related responses (Zhang and Luck, 2009) in humans. Such studies usually demonstrate enhanced responses to stimuli with features similar or equal to the attended feature, relative to stimuli with very different feature values (e.g., upward- vs downward-moving dots). Some even show a modulation of BOLD activity in completely unstimulated regions of the visual field due to selective feature-based attention (Serences and Boynton, 2007). Psychophysical experiments have also used perceptual aftereffects to measure the spatial spread of feature-based effects. For example, Boynton et al. (2006) instructed observers to selectively attend to one of two overlapping sets of moving dots on one side of the screen while ignoring another group of moving dots (the adaptor) in the opposite hemifield. The experimenters then measured the strength of the motion aftereffect (MAE) at the adaptor's location using dots moving very slowly either in the same or opposite direction of the previously attended dots. They found that the MAE was strongest when the adaptor had moved in the same direction as the attended dots. Additionally, a weak suppression effect on the MAE for the opposite motion direction was reported, albeit without a stable baseline to compare it against. These behavioral findings are supported by modulations of the BOLD signal from early visual cortex found in another study using a comparable paradigm (Saenz et al., 2002). Results of these experiments

were interpreted as evidence for the feature-similarity gain principle in humans.

These studies, along with most of the literature on feature-based attention, investigated how attention changes the magnitude of a response to a stimulus. However, the feature-similarity gain model might also imply that attention can shift a neural population's response in a way that the represented feature value is qualitatively changed as well. In a recent article in *The Journal of Neuroscience*, Zirnsak and Hamker (2010) set out to explore this key prediction of the feature-similarity gain model using a behavioral paradigm. As in Boynton et al. (2006), the underlying idea was to measure the motion aftereffect resulting from prolonged adaptation to an unattended moving stimulus, and to evaluate how feature-based attention to a second, distant stimulus affects that adaptation. However, rather than focus on the magnitude of the effect, Zirnsak and Hamker (2010) used the perceived direction of static test dots as an index of the direction tuning of the neurons that were most strongly adapted. In their study, observers looked for luminance changes in a group of moving dots (the target) on one side of the screen while ignoring a second group of moving dots (the adaptor) in the other hemifield. The target motion direction was either the same as or different from the adaptor. After about a minute, the adaptor dots were replaced with stationary dots, and subjects reported the direction of the resulting static motion aftereffect (SMAE).

Thus, the experimental logic was to measure the direction of the SMAE as a function of the direction of the attended target. The

Received May 28, 2010; revised June 30, 2010; accepted July 1, 2010.

The authors are supported by a Feodor-Lynen-Award from the Alexander von Humboldt Foundation (C.K.) and a graduate student fellowship from New York University (A.W.).

Correspondence should be addressed to Dr. Christian Kaul, Department of Psychology and Center for Neural Science, New York University, 6 Washington Place, New York, NY 10003. E-mail: ckaul@nyu.edu.

DOI:10.1523/JNEUROSCI.2777-10.2010

Copyright © 2010 the authors 0270-6474/10/3010261-02\$15.00/0

authors predicted that if feature-based attention truly biases the profile of activity in feature-specific neuronal populations across the visual field, the SMAE (measured on the unattended side) would change direction when the attended target's direction varies relative to the adaptor's direction. Such a change of direction of the MAE caused by feature-based attention has previously been established, albeit only when the attended and adapting stimuli were not spatially segregated (Alais and Blake, 1999).

Results indicated that when the target's motion direction was different from the adaptor's, the SMAE significantly changed direction compared with when the target's motion direction was identical with the adaptor's. For differences $<90^\circ$, the SMAE suggested that the representation of the adaptor's direction had been attracted toward the target's. Critically, this effect reversed for adaptor directions that were $>90^\circ$ from the target direction, as if the adaptor's direction had been repulsed. Thus, the direction of the SMAE depended on the angular difference between the directions of the adaptor and target. The authors explain these results with an extension of the feature-similarity gain model: of the neurons responding to the adaptor, those that prefer directions similar to the direction of the target were relatively boosted (and more strongly adapted), but those that prefer very different directions were relatively suppressed (and less strongly adapted). Thus, the peak of the population response to the adaptor shifted as a result of attention to the target. The direction of the SMAE (caused by this adaptor) then followed suit. In other words, the population response during adaptation resembled a pattern of activity that, without the influence of attention, would have been caused by an adaptor moving in a physically different direction.

In an improvement over previous behavioral work, Zirnsak and Hamker (2010) parametrically varied the target's direction, which allowed them to reveal a continuous spectrum of attraction and repulsion in feature space. But, as indicated in the methods, although the attraction effect was measured in four subjects, only two subjects were tested with the directions that showed the repulsion effect. Furthermore, in those two subjects, a slight imbalance in the design meant that fewer motion directions demonstrated the repulsion effect than showed the attraction effect. Given this small sample size, the repulsion effect on the SMAE should be confirmed in follow-up studies.

Zirnsak and Hamker (2010) demonstrate the effect of a neural principle derived from monkey electrophysiology (the

feature-similarity gain model) on human behavior. On the basis of their findings, they propose an extension of the existing model in which the influence of feature-based attention can distort the represented features of a stimulus. The authors discuss their result in the light of animal electrophysiology and previous models, but other human psychophysical and/or neuroimaging studies are also relevant. For example, Zirnsak and Hamker (2010) speculate that the origin of the modulatory influences of feature-based attention lies in V1. This may indeed be the case, but neuroimaging experiments have found effects of feature-based attention to motion direction in a range of visual areas, and not always in V1 (Serences and Boynton, 2007). It remains to be determined where in the brain these attentional signals originate, and in which visual areas they directly affect perception.

The study of Zirnsak and Hamker (2010) and other (human) studies summarized above (Saenz et al., 2002; Boynton et al., 2006; Serences and Boynton, 2007) attempt to measure the influence of feature-based attention on neurons with receptive fields distant from the attended stimulus. In those previous studies, subjects made perceptual judgments of one set of moving dots that were superimposed with distractor dots moving in a different direction. Thus, the subjects were required to selectively attend to a particular motion direction. Other work has argued that without a need for selective attention to filter out distractor features, there is no feature-specific spread of attentional effects (Saenz et al., 2003; Zhang and Luck, 2009). In Zirnsak and Hamker (2010), although subjects were asked to attend to the target's direction, motion was actually irrelevant to the primary task of detecting luminance changes in the target, as there was only one set of moving dots at that location. Accordingly, there is no independent measure of the effect of selective feature-based attention in Zirnsak and Hamker's primary task, and its role in the SMAE can only be inferred. These points raise the possibility that the mechanism hypothesized by Zirnsak and Hamker to explain their results is not identical to the type of feature-based attention that has previously been established in many behavioral, neuroimaging, and electrophysiology studies (Treue and Martínez Trujillo, 1999; Saenz et al., 2002; Boynton et al., 2006; Serences and Boynton, 2007).

Moreover, this lack of task requirement for attention to a specific feature reflects a general state of confusion in the literature about how to evoke feature-based attention. Is it really necessary for the observer to at-

tentively select the subset of stimuli moving in a particular direction, as in the human neuroimaging experiments (Saenz et al., 2002)? Or, is it sufficient for an attended object's motion to be task-relevant to engage the spread of feature-based attention (Martínez-Trujillo and Treue, 2004)? A third possibility, consistent with the results of Zirnsak and Hamker, is that all features of a task-relevant stimulus are automatically and globally boosted (Melcher et al., 2005). These potentially independent effects of attention should be carefully compared in future studies.

In conclusion, Zirnsak and Hamker have brought to light an important corollary to the feature-similarity gain model: that feature-based attention may distort the encoding of a moving object's direction. The exact nature of the mechanism at work could be clarified with a task that explicitly manipulates selective feature-based attention. As a final note, Zirnsak and Hamker measured the direction-shift phenomenon with an adaptation aftereffect, from which effects on processing of the previous adaptor were inferred. This raises a question for further study: was the perceived direction of the adaptor itself, a simultaneously present, distant stimulus, also distorted by attention?

References

- Alais D, Blake R (1999) Neural strength of visual attention gauged by motion adaptation. *Nat Neurosci* 2:1015–1018.
- Boynton GM, Ciaramitaro VM, Arman AC (2006) Effects of feature-based attention on the motion aftereffect at remote locations. *Vision Res* 46:2968–2976.
- Ling S, Liu T, Carrasco M (2009) How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Res* 49:1194–1204.
- Martínez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14:744–751.
- Melcher D, Papathomas TV, Vidnyánszky Z (2005) Implicit attentional selection of bound visual features. *Neuron* 46:723–729.
- Saenz M, Buracas GT, Boynton GM (2002) Global effects of feature-based attention in human visual cortex. *Nat Neurosci* 5:631–632.
- Saenz M, Buracas GT, Boynton GM (2003) Global feature-based attention for motion and color. *Vision Res* 43:629–637.
- Serences JT, Boynton GM (2007) Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55:301–312.
- Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–579.
- Zhang W, Luck SJ (2009) Feature-based attention modulates feedforward visual processing. *Nat Neurosci* 12:24–25.
- Zirnsak M, Hamker FH (2010) Attention alters feature space in motion processing. *J Neurosci* 30:6882–6890.