

Journal Club

Editor's Note: These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at www.jneurosci.org. For more information on the format, review process, and purpose of Journal Club articles, please see <http://jneurosci.org/content/preparing-manuscript>.

Facing a Regular World: How Spatial Object Structure Shapes Visual Processing

Daniel Kaiser¹ and Tristan Haselhuhn²

¹Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto, Italy, and ²Department of Educational Science, University of Regensburg, 93053 Regensburg, Germany

Review of de Haas et al.

In everyday environments, the visual system needs to process information from many stimuli appearing at different locations across the visual field. How can the brain optimally deal with the complexity of such inputs? Visual information processing in occipitotemporal regions of the visual cortex is separated for stimuli belonging to different categories, such as faces, bodies, scenes, or words (McCandliss et al., 2003; Downing et al., 2006). Most of these regions also contain retinotopic information and thus also separate stimulus processing with respect to location (Kravitz et al., 2013). Visual object analysis can therefore be understood as comprising specialized processing channels for distinct stimulus categories and for different retinotopic locations. As a result of this organization, when multiple stimuli need to be processed at the same time and there is a large degree of overlap in the processing channels (e.g., when stimuli are close together or stem from the same category), processing efficiency decreases. Conversely, a smaller degree of overlap can facilitate parallel processing of

visual information (Franconeri et al., 2013; Cohen et al., 2014).

Although many previous investigations suggest that location and identity information are processed through largely independent pathways, information about a stimulus' identity is rarely decoupled completely from its location (Kravitz et al., 2008). Indeed, identity and location information often interact as a consequence of real-world structure: clouds are usually above the head, whereas grass is underfoot. Experience with objects repeatedly appearing in specific locations of the visual field might shape processing channels that preferentially process these objects when they appear in their typical locations. Such channels would allow more efficient parallel processing of multiple stimuli and thus would provide a mechanism for optimal processing of complex, but regularly structured, visual information.

A recent paper in this journal (de Haas et al., 2016) provides evidence that stimulus identity and location interact during face perception. Faces contain a specific set of distinct parts (e.g., eyes, nose, mouth) that form a typical configuration (e.g., eyes above nose above mouth) and humans acquire massive experience with this highly predictable and often repeated configuration. de Haas et al. (2016) measured eye movement patterns to demonstrate that the typical relative locations of multiple face parts (e.g., the eyes appearing on both sides

and above the mouth) translate into typical absolute locations for single face parts (e.g., the eye more often appearing in the upper visual field and the mouth more often appearing in the lower visual field). Using fMRI, they showed that these location priors influence the processing of individual face parts presented in isolation, depending on the location they appear in: in the right inferior occipital gyrus, face parts appearing in their typical locations evoked more discriminable response patterns than face parts appearing in atypical locations. More specifically, the activity patterns evoked by viewing an eye versus viewing a mouth were more or less distinct depending on the retinotopic locations of the parts: the patterns were more discriminable when the eye was placed in the upper visual field and the mouth was placed in the lower visual field and less discriminable when the eye was placed in the lower visual field and the mouth was placed in the upper visual field. This difference in pattern discriminability indicates higher processing efficiency for typically positioned face parts: When there is less overlap in the mechanisms recruited for processing different face parts, information readout at later stages of the processing hierarchy is facilitated (Ritchie and Carlson, 2016). Thus, together with previous findings (Chan et al., 2010; Issa and DiCarlo, 2012), these results suggest that individual face parts appearing in their typical locations can be efficiently processed through

Received Nov. 7, 2016; revised Dec. 26, 2016; accepted Jan. 4, 2017.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Daniel Kaiser, Center for Mind/Brain Sciences, Corso Bettini 31, 38068 Rovereto, Italy. E-mail: danielkaiser.net@gmail.com.

DOI:10.1523/JNEUROSCI.3441-16.2017

Copyright © 2017 the authors 0270-6474/17/371965-03\$15.00/0

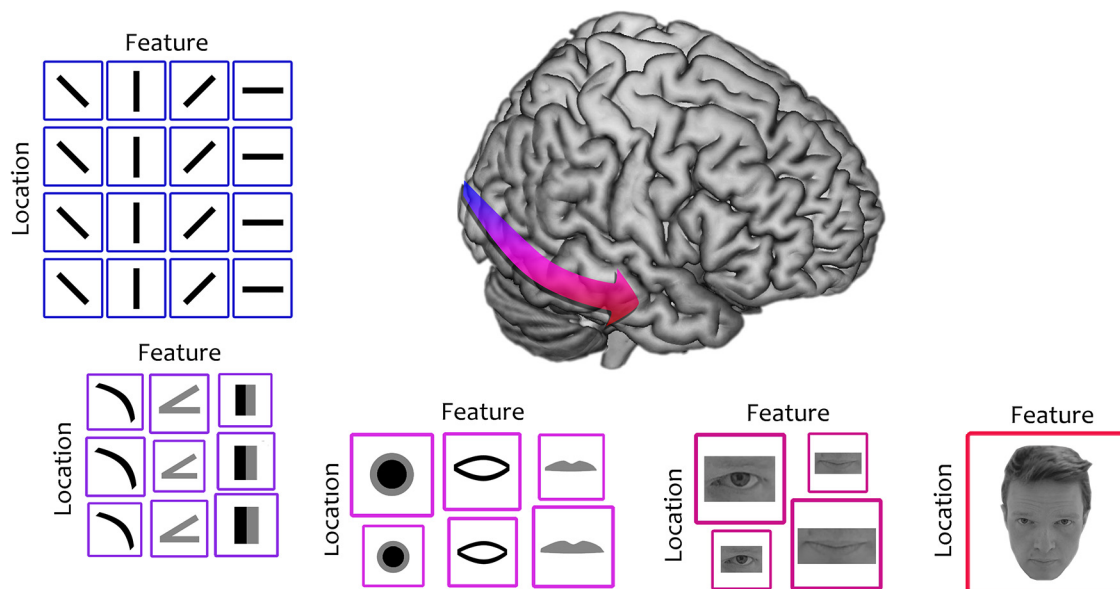


Figure 1. Location priors for face parts and their associated features can comprise neural processing channels that facilitate the processing of face parts in their typical locations. When moving along the ventral visual stream, receptive fields get bigger and preferred features get more complex. Because of their increasing complexity, features are increasingly likely to be linked to specific face parts (e.g., a dark disc surrounded by a lighter circle covaries with the presence of an eye). Information about such diagnostic visual features is fed into representations for specific face parts (e.g., an eye or a mouth), which in turn are fed into representations of whole faces later in the hierarchy. Because of repeated exposure, face-part-associated features could be preferentially processed when appearing in specific locations, where the associated part typically appears (exemplified by larger tiles). Further studies using time-resolved neural recordings are needed to investigate at which levels of the hierarchy location priors impact neural processing. Here, the location dimension reflects the vertical visual axis.

location-specific neural channels, reducing overlap in processing resources. Conversely, when face parts appear in atypical locations, they cannot readily be processed through these channels, leading to decreased processing efficiency. These findings suggest that face processing in visual cortex can be partly explained by neural channels optimally tuned for individual face parts appearing in their typical locations.

Which levels of the visual hierarchy do these processing channels span? One possibility is that the vast experience with specific face parts appearing in specific spatial locations shapes processing for face parts and their associated visual features alike (Fig. 1). On this account, visual features that are characteristic for a specific face part (e.g., the outline of an eye or the characteristic contrast between iris and pupil) are repeatedly processed through specific retinotopic neural populations. This repeated processing is expected to enhance neural efficiency in coding such features, relative to other neural populations that more rarely process the same features. Supporting this hypothesis, data from nonhuman primates revealed a relatively early benefit of typical (vs atypical) retinotopic positioning of the eye (Issa and DiCarlo, 2012). Within the first 100 ms of processing, cells in face-selective regions of the inferotemporal cortex respond more strongly to eyes appearing in the contralateral upper visual field than to eyes appearing in other

retinotopic locations, suggesting that already the processing of more basic, stimulus-associated visual features is affected (Carlson et al., 2013; Kaiser et al., 2016). Neural processing channels connecting such location-specific feature representations and location-specific face-part representations could support optimal visual analysis across multiple levels. However, further studies in humans are needed to investigate how profoundly location priors for face parts influence visual processing at different stages. These studies could exploit recent advances in multivariate decoding of magnetoencephalography data (Carlson et al., 2013; Cichy et al., 2014) to reveal whether location priors affect neural responses throughout the visual hierarchy, potentially starting from more basic features that are associated with specific face parts.

Importantly, de Haas et al. (2016) demonstrate that the neural processing advantage for typically positioned face parts is reflected in more accurate perception. In a behavioral experiment, individual face parts were more accurately discriminated when appearing in their typical locations (e.g., an eye appearing in the upper visual field) than when they appeared in atypical locations (e.g., an eye appearing in the lower visual field). This finding is exciting because spatial structure might similarly shape efficient processing channels for other types of visual input, helping to explain behavioral perfor-

mance when inputs are complex but contain spatial regularities.

One example of such inputs is written text. Words are composed of single letters, which do not appear in random order, but are arranged in meaningful chunks. Activity in visual cortex has been shown to reflect the regularity structure within groups of letters, revealing a tuning for frequently experienced letter combinations (Vinckier et al., 2007). Could neural responses also be sensitive to the absolute, retinotopic locations of individual text elements? Often letters (or combinations of letters) are found in specific positions within a word: Some are more likely to appear at the beginning of words (e.g., the letter “j” or the prefix “pro”), whereas others are more likely to appear at the end of words (e.g., the letter “y” or the suffix “ing”). During reading, each word is typically targeted only by one or two fixations, which most frequently fall close to the middle of the word (Rayner, 1979; Clifton et al., 2016). As a consequence of such eye movement patterns, the recurring relative positioning of letters could introduce location priors for single letters or chunks of letters (e.g., “j” appearing more often to the left of a fixation and “y” more often to the right). The processing of different text elements should thus not be uniformly pronounced across the visual field. Conversely, each letter (or letter combination) is likely to have a preferred position in the visual field depending on where it most frequently ap-

pears relative to fixation. Text elements that adhere to these typical locations could be more efficiently processed on a neural level, offering a new perspective for studying reading performance.

Another example is the processing of objects in the context of real-world environments. Natural scenes typically consist of numerous distinct objects that are arranged in a meaningful way. Constrained by their functions and physical constraints, objects appear in highly predictable locations within the environment. Previous work has shown increased processing efficiency for typically (vs atypically) positioned multiobject arrangements of objects in ventral visual cortex (Kaiser et al., 2014). Similarly to the regular structure among face parts, such regularities among multiple objects (e.g., lamps appearing above tables) could translate into typical retinotopic positions for single objects (e.g., lamps appearing in the upper visual field). Such location priors for single objects could enable enhanced neural processing of single objects when they appear in their typical, retinotopic locations. Location-specific processing channels for individual objects within complex scenes would provide a novel explanation for the striking efficiency of natural scene perception compared with simple stimulus displays, where no location priors for individual display elements exist (Peelen and Kastner, 2014).

In conclusion, neural processing channels that integrate stimulus identity and location might not only enable the brain to efficiently process the multiple parts of

a face. The work by de Haas et al. (2016) could highlight a general mechanism supporting efficient neural coding and perception of various types of complex, but regular, visual input.

References

- Carlson T, Tovar DA, Alink A, Kriegeskorte N (2013) Representational dynamics of object vision: the first 1000 ms. *J Vis* 13:10. [CrossRef Medline](#)
- Chan AW, Kravitz DJ, Truong S, Arizpe J, Baker CI (2010) Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nat Neurosci* 13:417–418. [CrossRef Medline](#)
- Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. *Nat Neurosci* 17:455–462. [CrossRef Medline](#)
- Clifton C, Ferreira F, Henderson JM, Inhoff AW, Liversedge SP, Reichle ED, Schotter ER (2016) Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *J Mem Lang* 86:1–19. [CrossRef](#)
- Cohen MA, Konkle T, Rhee JY, Nakayama K, Alvarez GA (2014) Processing of multiple visual objects is limited by overlap in neural channels. *Proc Natl Acad Sci U S A* 111:8955–8960. [CrossRef Medline](#)
- de Haas B, Schwarzkopf DS, Alvarez I, Lawson RP, Henriksson L, Kriegeskorte N, Rees G (2016) Perception and processing of faces in the human brain is tuned to typical facial feature locations. *J Neurosci* 36:9289–9302. [CrossRef Medline](#)
- Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N (2006) Domain specificity in visual cortex. *Cereb Cortex* 16:1453–1461. [CrossRef Medline](#)
- Franconeri SL, Alvarez GA, Cavanagh P (2013) Flexible cognitive resources: competitive content maps for attention and memory. *Trends Cogn Sci* 17:134–141. [CrossRef Medline](#)
- Issa EB, DiCarlo JJ (2012) Precedence of the eye region in neural processing of faces. *J Neurosci* 32:16666–16682. [CrossRef Medline](#)
- Kaiser D, Stein T, Peelen MV (2014) Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc Natl Acad Sci U S A* 111:11217–11222. [CrossRef Medline](#)
- Kaiser D, Azzalini DC, Peelen MV (2016) Shape-independent object category responses revealed by MEG and fMRI decoding. *J Neurophysiol* 115:2246–2250. [CrossRef Medline](#)
- Kravitz DJ, Vinson LD, Baker CI (2008) How position dependent is visual object recognition? *Trends Cogn Sci* 12:114–122. [CrossRef Medline](#)
- Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M (2013) The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci* 17:26–49. [CrossRef Medline](#)
- McCandliss BD, Cohen L, Dehaene S (2003) The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn Sci* 7:293–299. [CrossRef Medline](#)
- Peelen MV, Kastner S (2014) Attention in the real world: toward understanding its neural basis. *Trends Cogn Sci* 18:242–250. [CrossRef Medline](#)
- Rayner K (1979) Eye guidance in reading: fixation locations within words. *Perception* 8:21–30. [CrossRef Medline](#)
- Ritchie JB, Carlson TA (2016) Neural decoding and “inner” psychophysics: a distance-to-bound approach for linking mind, brain, and behavior. *Front Neurosci* 10:190. [CrossRef Medline](#)
- Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L (2007) Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron* 55:143–156. [CrossRef Medline](#)