

Understanding mid-level representations in visual processing

Jonathan W. Peirce

Nottingham Visual Neuroscience, School of Psychology,
University of Nottingham, Nottingham, UK



It is clear that early visual processing provides an image-based representation of the visual scene: Neurons in Striate cortex (V1) encode nothing about the meaning of a scene, but they do provide a great deal of information about the image features within it. The mechanisms of these “low-level” visual processes are relatively well understood. We can construct plausible models for how neurons, up to and including those in V1, build their representations from preceding inputs down to the level of photoreceptors. It is also clear that at some point we have a semantic, “high-level” representation of the visual scene because we can describe verbally the objects that we are viewing and their meaning to us. A huge number of studies are examining these “high-level” visual processes each year. Less well studied are the processes of “mid-level” vision, which presumably provide the bridge between these “low-level” representations of edges, colors, and lights and the “high-level” semantic representations of objects, faces, and scenes. This article and the special issue of papers in which it is published consider the nature of “mid-level” visual processing and some of the reasons why we might not have made as much progress in this domain as we would like.

On the need to study mid-level vision

It seems very clear that, at the earliest levels of visual processing, we have a number of image-based representations of the visual scene. For example, at the level of photoreceptors we have a representation of the distribution of light across the retina. At the level of retinal ganglion cells we have a representation of the changes in light (spatiotemporal contrast) across the retina due to combinations of excitatory and inhibitory inputs (Hartline, 1938; Kuffler, 1953). In these representations the responses of individual units (e.g., photoreceptors, ganglion cells) are driven largely by the pattern of light in the visual scene, with little dependence on what in the physical world gave rise to that pattern of light. We know a great deal about these

early stages of visual processing: how the light is detected; how contrast, orientation, spatial frequency, and color are represented; and how neurons change their responses over time (for a primer see Lennie, 2003). We can also use models constructed from one type of stimulus, such as gratings, to explain a fair amount of variance in responses to completely different stimuli, such as naturalistic movies (Mante, Bonin, & Carandini, 2008). By the level of the primary visual cortex, this image-based representation provides all the information needed for the visual system to identify simple visual features, such as edges, and encode the characteristics of those simple features.

It is also clear that at some point we must also have some form of semantic representation. Human observers can trivially describe the semantic content of an image; they know that this contains a face or a chair and can describe the characteristics of the objects while ignoring the attributes of the image per se. This higher level semantic representation of the scene is not as well understood as low-level visual processing—we do not have physiologically based computer models of face recognition, for instance—but these representations have certainly received a great deal of attention that has informed us of their general characteristics. We know, for instance, that a region in the fusiform gyrus responds significantly more strongly to the presentation of face stimuli than to others (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997) and that this representation is sensitive to the orientation of the face (Kanwisher, Tong, & Nakayama, 1998) but is relatively insensitive to its retinal location (Andrews & Ewbank, 2004).

We know very little about the intervening processes. Indeed, it is not necessarily the case that any intervening levels of processing exist. All of the information required to build the high-level semantic representation must already exist in the image-based representation of Striate cortex (V1) (or earlier); from these representations no new information is added—only transformed. It might be, therefore, that the information used in a high-level semantic representation is retrieved directly from V1 and/or earlier areas.

Citation: Peirce, J. W. (2015). Understanding mid-level representations in visual processing. *Journal of Vision*, 15(7):5, 1–9, doi:10.1167/15.7.5.

Search term	Articles (<i>N</i>)
(V1 [Title] or <i>striate</i> [Title] or <i>primary visual cortex</i> [Title]) and <i>visual</i>	2,725
V2 [Title] and <i>visual</i>	238
V3 [Title] and <i>visual</i>	37
V4 [Title] and <i>visual</i>	229

Table 1. Comparison of National Center for Biotechnology Information PubMed searches for studies studying primarily one or more retinotopic visual cortices (retrieved March 2015). *Notes:* For each search, the name of the visual area was required in the title and the word *visual* was required somewhere in the document (because the terms *V1* and *V2* are often used to refer to version numbers). This search is meant to be illustrative rather than exhaustive.

However, it might be that there are one or more intervening stages with increasingly complex representations of the visual scene, continuing in the manner seen in the early stages of visual processing.

As well as there being, potentially, several stages of visual processing that we have not uncovered, there are several anatomical regions whose contribution to visual processing we have largely failed to understand. Comparatively few laboratories so far have grappled with cortical areas V2, V3, and V4. One suggestion for V2 is that it might be involved in detecting contours in the absence of net change in chromaticity or luminance (von der Heydt, Peterhans, & Baumgartner, 1984), although this might occur as early as V1 (Grosf, Shapley, & Hawken, 1993). It has also been suggested that this area is sensitive to border ownership (e.g., Zhou, Friedman, & von der Heydt, 2000). Most recently, V2 has been found to be sensitive to higher order statistics in visual patterns (Freeman, Ziemba, Heeger, Simoncelli, & Movshon, 2013).

V4 was originally proposed as a center for the coding of color (e.g., Zeki, 1973), although subsequent studies have questioned its special role in that domain (e.g., Krüger & Gouras, 1980; Schein, Marrocco, & de Monasterio, 1982). More recently it has been associated with curvature perception (e.g., Pasupathy & Connor, 1999, 2002; Yau, Pasupathy, Brincat, & Connor, 2013). For V2 and V4, then, it is not that there have been no studies at all, but rather very few have been conducted compared with the number of laboratories studying low-level areas, such as V1 (see Table 1), or those involved in high-level processing, such as fusiform face area. The rate of progress in understanding V2 and V4 is, naturally, slowed by this paucity of laboratories trying to understand them.

There are multiple visual cortical areas about which we know relatively little compared with low- and high-level visual regions, and there is a transform from an image-based representation of edges to a semantic representation of objects that we have yet to under-

stand. Understanding these is surely one of the holy grails in understanding visual processing—at least, in understanding visual form perception (the ventral pathway of Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). So why are so few vision scientists trying to understand the intermediate representations of visual processing?

While presenting my laboratory's studies to the vision science community, the objections I have heard have fallen into three categories, and I think that these are likely to be more generally about mid-level vision than about my attempts to study it. The first broad camp of opposition is that these intermediate representations obviously exist and therefore they are not worth studying. The second camp of vision scientists says that these intermediate representations could not possibly exist and so we should not try to study them. The third form of opposition is that intermediate representations probably do exist but we could never understand them, or at least describe them, so they probably are not worth studying. It seems that the community is, at least, in agreement that the mechanisms are not worth studying, but let us consider more closely the reasoning why.

Let us consider the possibility of mid-level detectors with respect to a putative mechanism for encoding curvature. Is it obvious that intermediate mechanisms for a curvature detector must exist? Their existence is certainly assumed, or implicit, to many theories and studies. To some computational models they are explicit (e.g., Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; Wilson, Wilkinson, & Asaad, 1997). However, all the information required to construct a representation of curvature is actually available from the preceding (established) representation of orientation and space. There is technically no need for curvature ever to be formally represented as an entity. High-level processing mechanisms could potentially retrieve that information directly from the preceding level and skip the intervening step. Therefore, it does not seem a certainty that intermediate representations exist.

The main reason for the second objection—that intermediate representations obviously could not exist—is that simply combining signals from each layer to make more complex representations would require far too many computational units (e.g., neurons). For instance, if we were to combine in a pairwise fashion the outputs of all V1 neurons in area V2, then V2 would need to contain the square of the number of neurons in V1, which it clearly does not. Similarly, the next stage would need some factorial combination of the cells in V2, and we quickly find a combinatorial explosion of entities that need to be encoded. This is clearly a problem that needs to be considered and addressed, but maybe not a reason to discard the entire

endeavor. The circuits leading up to the primary visual cortex certainly are taking essentially this approach. The representation of spatiotemporal contrast found in ganglion and lateral geniculate nucleus (LGN) cells is constructed by a center-surround organization of excitatory and inhibitory inputs from photoreceptors via a complex intermediate circuit involving horizontal, bipolar, and amacrine cells (Dowling & Boycott, 1966). Similarly, by combining the outputs of these units we could create oriented receptive fields with inhibitory and excitatory regions, as found in V1 simple cells (Hubel & Wiesel, 1962). People do not seem to have raised the combinatorial explosion issue for these steps; there is not a claim that there are too many possible combinations for V1 to represent combined LGN outputs. Possibly, that is because V1 has more neurons than the LGN, but note that V1 contains approximately 140 million neurons (Wandell, 1995) compared with the approximately 1.5 million retinal ganglion cells (Hecht, 2001), an increase that is insufficient to handle the sort of exponential increase implied by the combinatorial explosion. Furthermore, the step from photoreceptor to ganglion cell constitutes a reduction of processing units from approximately 4.6 million (Curcio, Sloan, Kalina, & Hendrickson, 1990) to approximately 1.5 million retinal ganglion cells (Hecht, 2001). It seems curious, then, that we are willing to accept that LGN outputs are combined by V1 to make a new, more complex representation but unwilling to accept that some similar process occurs beyond that. Indeed, given that neurons work by combining, with a mixture of excitation and inhibition, the outputs of the preceding layer (and potentially feedback from higher layers), it seems that some form of combination of outputs must occur. The problem, then, is “simply” to understand what form of combination that takes: what the representation looks like, how many units’ signals are combined in any one step, and how the combinatorial explosion is averted.

The third objection that I have encountered is that, although some form of signal combination might occur, we could never understand or express its representation using our human intuition and language. If the brain represents the content of a visual scene not by individual neurons’ encoding dimensions that we understand but by the state of its networks, then the combinatorial problem disappears. A simplified model network with only 100 neurons and 10 levels of available response for each neuron obviously has 10^{100} available states, which means it can represent a phenomenal number of things. This is true, however, only if we expect any individual unit’s response to be essentially arbitrary, not conforming to our human preconceptions of what might be “meaningful.” As we impose meaning (i.e., constraints) on the responses of individual units (e.g., requiring that neighboring

processing units respond to neighboring parts of visual space or that contrast response curves are smooth functions), the number of valid unique states decreases.

In multilayer artificial neural nets, trained to discriminate between a range of stimuli, it is common to have an *input* layer, and *output* layer, and at least one *hidden* layer (see Rosenblatt, 1962, for a detailed introduction to multilayer networks). After training, our artificial neural net might well perform successful discriminations—the hidden layer may have created representations of the input that can be classified by the output layer—but the responses of the hidden layer are not necessarily meaningful or understandable to us when we inspect them. The responses of individual units in this layer have no reason to find the features that we, as humans, intuitively appreciate in the stimuli.

If the units are not responding to features of the world that appear to us to have meaning, then the search for intermediate representations might be extremely difficult. Indeed, that might explain why we have made relatively little progress in this area of visual processing. However, the strong form of the hypothesis—that there is no meaning to the responses of a single unit in the intermediate network layers—is certainly incorrect. The fact that the visual cortices are retinotopically organized, for instance, means that the responses of a neuron in, say, V3 are not entirely arbitrary because they are based on a relatively localized area of the visual scene. The fact that neighboring neurons relate to neighboring regions of space is a sign that there is organization to the visual code in these cortices and that the scene is not encoded as an arbitrary distributed code as the hidden layers of an artificial neural network might be.

It seems clear, then, that intermediate mechanisms might exist and that understanding the code of these mechanisms might be a useful endeavor. This special issue of *Journal of Vision* considers a number of ways in which the mechanisms might be investigated using a variety of methods from the neuroscientist’s toolkit, such as psychophysics (Loffler, 2015; Wilson & Wilkinson, 2015) and neuroimaging (Andrews, Watson, Rice, & Hartley, 2015; Kourtzi & Welchman, 2015). The aim of the articles, and of the Vision Science Society Symposium on which they were based (Andrews, 2014; Kourtzi, 2014; Loffler, 2014; Pasupathy, 2014; Peirce, 2014; Wilson, 2014), is to highlight the range of ways in which we can address this most complex of problems.

Selective mechanisms for simple conjunctions: Plaids

For the remainder of this article we consider the potential units that my laboratory has been investi-

gating, representing simple combinations of gratings that form plaid patterns and simple, short, curved contours. Studying these seemed to be a natural extension of the logic by which encoding appeared to progress in earlier levels of processing. From selective combinations (both excitatory and inhibitory) of signals from photoreceptors, we can see how the receptive fields of retinal ganglion cells (and lateral geniculate nucleus neurons) might operate (Dowling & Boycott, 1966). From selective combinations of those, again in an excitatory and inhibitory fashion, we can create plausible models of how a V1 simple cell might work (Hubel & Wiesel, 1962), and we can use combinations of those to produce a spatiotemporal energy model (Adelson & Bergen, 1985) to simulate complex cells. Indeed, the nature of neurons is that they essentially add and subtract signals from a set of inputs, albeit in variously nonlinear ways. That being the case, we might wonder what the next form of combination might be given the well-understood spatiotemporal receptive fields of V1 neurons.

This leads us to a testable hypothesis; we could search for evidence of neural and perceptual mechanisms responding selectively to particular combinations of sinusoidal gratings. My laboratory has concentrated on the particular combinations that form plaid patterns (i.e., a pair of oriented components with substantial spatial overlap and dissimilar orientations) and curved contours (with minimal spatial overlap and similar orientations). These are not the only forms of conjunction, of course; they are merely simple combinations of pairs of gratings.

There is some reason to think that there might be special mechanisms for plaid patterns from the fact that some plaids, but not all, appear perceptually as coherent patterns (Adelson & Movshon, 1982; Meese & Freeman, 1995). When high-contrast orthogonal gratings are combined to form a plaid pattern and are then drifted behind an aperture, the percept is that of a single surface moving in one direction. On the other hand, if the gratings differ in spatial frequency or contrast patterns or if the angle between the components is very oblique, then we perceive two semitransparent surfaces moving past each other (Adelson & Movshon, 1982). Similarly, static plaids can be perceived as coherent, unitary checkerboard patterns or as two perceptually distinct gratings (Meese & Freeman, 1995). In Figure 1, for example, in the upper row the plaid patterns all look like checkerboards and one struggles to perceive the gratings from which they are constructed. In the lower row there is an increasing difference in spatial frequencies, causing the percept of a coherent plaid gradually to break. In the lower right pattern the spatial frequency (SF) of one component is four times greater than that of the other, and the percept is clearly of two separate gratings.

In an attempt to identify selective mechanisms for plaids more directly, we conducted a number of studies measuring the selective contrast adaptation to a plaid pattern—the extent to which the perceived contrast of a plaid decreases after prolonged exposure. This was compared with the degree of adaptation to the same grating components combined into other plaid patterns (Peirce & Taylor, 2006) or presented for equivalent adaptation periods as single component gratings (Hancock, McGovern, & Peirce, 2010; McGovern, Hancock, & Peirce, 2011; McGovern & Peirce, 2010). The results showed consistently greater adaptation to the plaids than predicted by adaptation to the gratings alone, provided that the contrast of the probe stimuli was reasonably high (McGovern & Peirce, 2010) and, as with the perceptual reports above, that the SFs of the gratings were similar (Hancock et al., 2010). The findings are consistent with a mechanism responding selectively to the presence of a plaid pattern, which we conceive as a unit performing a logical AND operation on the outputs of two SF channels.

We are not the only group to have suggested such mechanisms. Robinson and MacLeod (2011) also used adaptation with compound gratings to show a version of the McCollough effect with plaid patterns. Nam, Solomon, Morgan, Wright, and Chubb (2009) showed pop-out effects for plaids in a visual search task, which they put down to a preattentive mechanism for which the whole is greater than the sum of the parts. In their study the relative SFs of the components were also manipulated and, again, the finding was that the effect of visual pop-out was reduced if the SFs were dissimilar (Nam et al., 2009).

Selective mechanisms for simple conjunctions: Curvature

Plaids are obviously only one form of conjunction that we can create with a pair of sinusoidal gratings. We also considered whether signals of small groups of “edges” would be combined by similar selective mechanisms for curvature. Previous groups have studied the sensitivity of trained observers in detecting patterns of Gabor elements in paths in a contour integration task (Field, Hayes, & Hess, 1993). Elements can be detected in either the *snake* or the *ladder* configuration of the contour in which the elements are aligned to, or orthogonal with, the orientation of the path (Bex, Simmers, & Dakin, 2001; Field et al., 1993; Ledgeway, Hess, & Geisler, 2005). A variety of *association field* models have been developed to model the conditions where we are more and less sensitive to the presence of contours in these and radial frequency patterns (quasicircular patterns that are typically

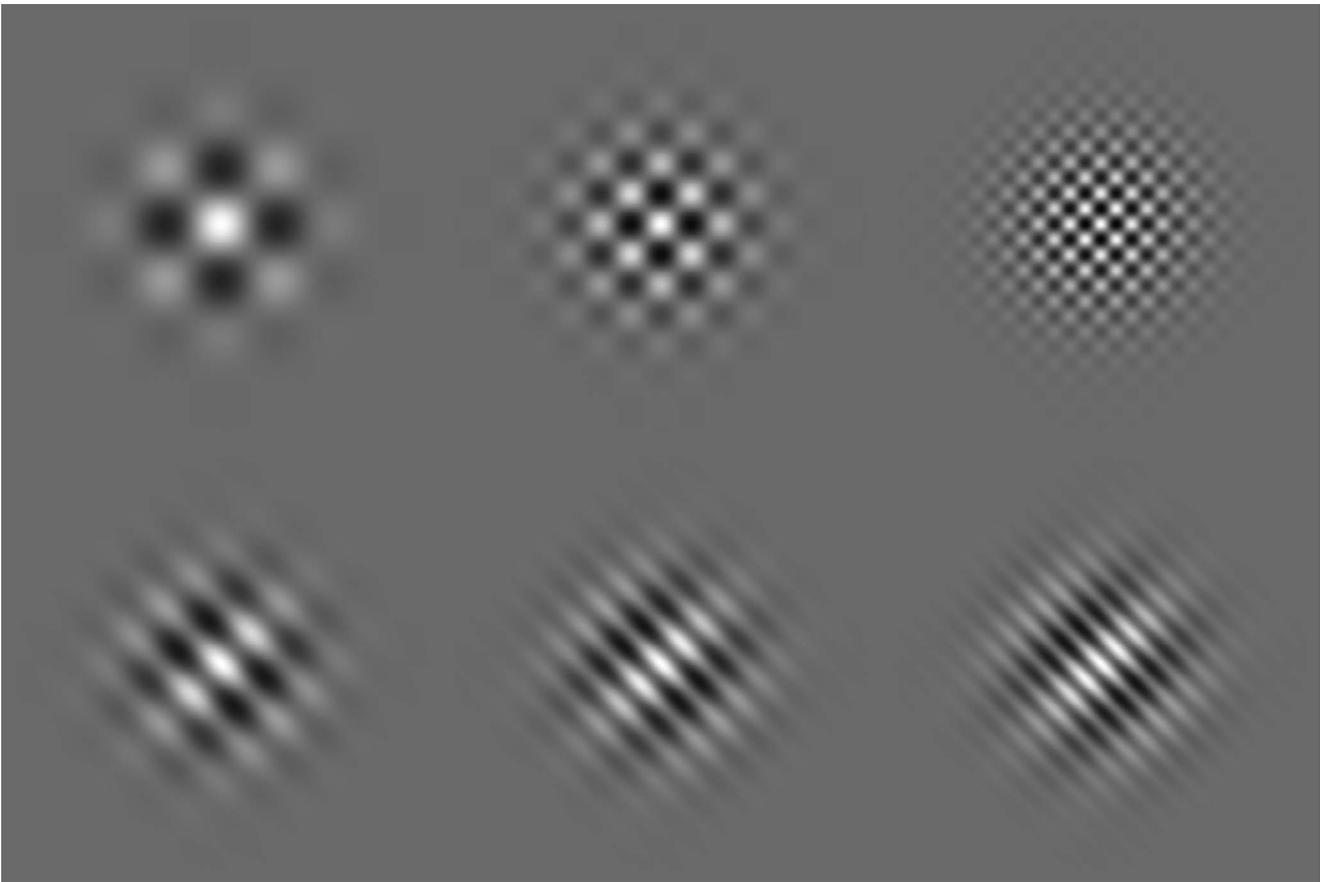


Figure 1. When high-contrast orthogonal gratings matched in SF are combined, the percept is always of a single checkerboard pattern (top row) rather than two gratings. As SFs become increasingly disparate (bottom row component SFs differ by two, three, and four times, respectively), it becomes easy to identify the two component gratings, and the percept is less clearly a plaid pattern.

distinguished from truly circular patterns). Similarly, measurements of our sensitivity to the global form of arrays of elements (reviewed in this issue by Wilson & Wilkinson, 2015) and our sensitivity to radial frequency patterns (reviewed in this issue by Loffler, 2015; Wilson & Wilkinson, 2015) have all suggested specific perceptual mechanisms for the processing of curvature in form perception.

As with the plaid studies, we might also use adaptation to try to identify the extent to measure and characterize mechanisms sensitive to conjunctions. We reasoned that we might be able to design a method of compound adaptation for curvature similar to the method we used in plaid adaptation (Hancock & Peirce, 2008). We extended the idea of the tilt aftereffect (Gibson & Radner, 1937) to a curvature aftereffect by adapting subjects to a pair of Gabor patches with slight overlap and slightly different orientations that combine perceptually to form a curved contour (or chevron, if the orientation difference is substantial). As in the plaid adaptation experiments, we were careful to equate the local tilt aftereffects caused by the components by having a

second field in which the same components were presented as adapters for the same period of time but alternated with each other so that the curved contour itself was not present. In both the component-adapted field and the compound-adapted field a straight contour then appears to curve in the opposite direction, but, notably, a greater repulsion from the adapting stimulus is found in the compound-adapted field (Hancock & Peirce, 2008), as expected if we have adapting curvature-selective visual mechanisms. Similar conclusions come from the findings of shape-frequency and shape-amplitude aftereffects by Gheorghiu and Kingdom (2006, 2007, 2008), although the complexity of this effect means that many more factors need to be controlled (Gheorghiu & Kingdom, 2008).

Another form of more global adaptation was discovered recently that might also be taken as evidence for perceptual mechanisms for curvature: adaptation to the tilt of an edge that is *implied* by a context. If participants are adapted to a circular pattern with a missing section then, when a standard linear grating is used as a probe in the missing section, it shows a form

of tilt aftereffect (Roach & Webb, 2013; Roach, Webb, & McGraw, 2008). Unlike the tilt aftereffect, however, this effect is untuned to SF differences between adapter and probe (Roach et al., 2008) and its effects are maintained over large spatial separations between them (Roach & Webb, 2013). Furthermore, it works only for radial and concentric inducers; a traditional linear grating adapter does not generate an implied tilt effect in the same way (Roach et al., 2008).

Selective mechanisms for simple conjunctions: The parameter explosion

If the mechanisms suggested here were truly to exist, we would need to find a solution for the problem of parameter explosion described earlier. The visual system cannot simply represent all possible combinations of gratings in a pairwise (or worse) manner.

There are two basic ways to get around this problem: increasing *selectivity* about which components you combine, and increasing *invariance* to parameters that are no longer of interest at the new level of representation. We have seen evidence of the former strategy in the fact that plaids are perceived as such only if the components are similar in SF or other parameters (Adelson & Movshon, 1982; Hancock et al., 2010; Meese & Freeman, 1995). Why that would be the case is unclear at this point; we do not know why certain combinations of gratings are perceived as a unitary pattern while others are not. We might speculate that some have greater prevalence or relevance in natural scenes. We found the same selectivity effect for the curvature aftereffect (Hancock et al., 2010), and in this case the reason seems obvious: If the SFs of the components are dissimilar, then they probably do not form a contour originating from a single visual object.

Introducing invariance to a particular stimulus dimension seems the more potent method of reducing the space that an encoding mechanism needs to capture. The only test for invariance that we carried out in the plaid and curvature compound adaptation studies was using SF and, for that, we found no sign of invariance. When the adapter and probe differed from each other in SF (but with components that were matched), the selective adaptation disappeared, which does not seem indicative of SF-invariant tuning of either mechanism (Hancock et al., 2010). Of course, there are other dimensions over which gratings can vary that we did not test, mostly for practical reasons. For example, we briefly investigated using chromatic gratings and combinations in which the chromaticity

varied between components, which also prevents binding of the components into a coherent contour. When this did not yield significant selective adaptation, we could not be certain that the reason was a lack of conjunction processing or simply weaker adaptation to the components (chromatic gratings necessarily have a lower effective contrast on standard displays).

Unlike the curvature aftereffect that we found, the implied tilt effect of Roach et al. (2008) does appear to be untuned to SF. Presumably, the mechanism that supports this is at a later stage of processing than the mechanism for the simple curvature aftereffect that Hancock and Peirce (2008) measured.

Summary

Many vision scientists consider mid-level mechanisms for coding of conjunctions to be so obvious that we should not study them. Others consider these mechanisms so obviously implausible that we should not study them. Of the remaining vision scientists, several consider these mechanisms so hard to study that we should not study them. However, an increasing number of us are trying to find ways to expose the mechanisms falling between low-level edge detection and high-level object perception. Given the rich variety of techniques available to modern vision science, all of which will be needed for this seminal question, this seems an ideal time to try to understand the representations of mid-level vision.

Keywords: mid-level vision, plaids, curvature, hierarchy, V2, V3, V4

Acknowledgments

This work was supported by grants from the BBSRC (BB/C50289X/1) and the Wellcome Trust (WT085444). Many thanks to all my colleagues in the Nottingham Visual Neuroscience group for their feedback on these ideas, and especially to David McGovern and Sarah Hancock who together collected most of the data on compound adaptation methods described in this review.

Commercial relationships: none.

Corresponding author: Jonathan W. Peirce.

Email: jonathan.peirce@nottingham.ac.uk.

Address: Nottingham Visual Neuroscience, School of Psychology, University of Nottingham, Nottingham, UK.

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*, 284–299, doi:10.1364/JOSAA.2.000284.
- Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*, 523–525.
- Andrews, T. J. (2014). Low-level image properties of visual objects explain category-selective patterns of neural response across the ventral visual pathway. *Journal of Vision*, *14*(10):1459, doi:10.1167/14.10.1459. [Abstract]
- Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*, *23*, 905–913, doi:10.1016/j.neuroimage.2004.07.060.
- Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision*, *15*(7):3, 1–12, doi:10.1167/15.7.3. [PubMed] [Article]
- Bex, P. J., Simmers, A. J., & Dakin, S. C. (2001). Snakes and ladders: The role of temporal modulation in visual contour integration. *Vision Research*, *41*, 3775–3782.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, *292*, 497–523, doi:10.1002/cne.902920402.
- Dowling, J. E., & Boycott, B. B. (1966). Organization of the primate retina: Electron microscopy. *Proceedings of the Royal Society of London B*, *166*, 80–111, doi:10.1098/rspb.1966.0086.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local “association field.” *Vision Research*, *33*, 173–193.
- Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*, 974–981, doi:10.1038/nn.3402.
- Gheorghiu, E., & Kingdom, F. A. A. (2006). Luminance-contrast properties of contour-shape processing revealed through the shape-frequency after-effect. *Vision Research*, *46*, 3603–3615, doi:10.1016/j.visres.2006.04.021.
- Gheorghiu, E., & Kingdom, F. A. A. (2007). The spatial feature underlying the shape-frequency and shape-amplitude after-effects. *Vision Research*, *47*, 834–844, doi:10.1016/j.visres.2006.11.023.
- Gheorghiu, E., & Kingdom, F. A. A. (2008). Spatial properties of curvature-encoding mechanisms revealed through the shape-frequency and shape-amplitude after-effects. *Vision Research*, *48*, 1107–1124, doi:10.1016/j.visres.2008.02.002.
- Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, *20*, 453–467, doi:10.1037/h0059826.
- Goodale, M., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25, doi:10.1016/0166-2236(92)90344-8.
- Grosf, D. H., Shapley, R. M., & Hawken, M. J. (1993). Macaque V1 neurons can signal “illusory” contours. *Nature*, *365*, 550–552, doi:10.1038/365550a0.
- Hancock, S., McGovern, D. P., & Peirce, J. W. (2010). Ameliorating the combinatorial explosion with spatial frequency-matched combinations of V1 outputs. *Journal of Vision*, *10*(8):7, 1–14, doi:10.1167/10.8.7. [PubMed] [Article]
- Hancock, S., & Peirce, J. W. (2008). Selective mechanisms for simple contours revealed by compound adaptation. *Journal of Vision*, *8*(7):11, 1–10, doi:10.1167/8.7.11. [PubMed] [Article]
- Hartline, H. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology*, *121*, 400–415, doi:10.1234/12345678.
- Hecht, E. (2001). *Optics* (4th ed., Vol. 1). Reading, MA: Addison-Wesley.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, *160*, 106–154, doi:10.1523/JNEUROSCI.1991-09.2009.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*, 4302–4311.
- Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, *68*, B1–B11, doi:10.1016/S0010-0277(98)00035-3.
- Kourtzi, Z. (2014). Adaptive shape coding in the human visual brain. *Journal of Vision*, *14*(10):1457, doi:10.1167/14.10.1457. [Abstract]

- Kourtzi, Z., & Welchman, A. E. (2015). Adaptive shape coding for perceptual decisions in the human brain. *Journal of Vision*, *15*(7):2, 1–9, doi:10.1167/15.7.2. [PubMed] [Article]
- Krüger, J., & Gouras, P. (1980). Spectral selectivity of cells and its dependence on slit length in monkey visual cortex. *Journal of Neurophysiology*, *43*, 1055–1069.
- Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*, 37–68.
- Ledgeway, T., Hess, R., & Geisler, W. (2005). Grouping local orientation and direction signals to extract spatial contours: Empirical tests of “association field” models of contour integration. *Vision Research*, *45*, 2511–2522, doi:10.1016/j.visres.2005.04.002.
- Lennie, P. (2003). Receptive fields. *Current Biology*, *13*, R216–R219, doi:10.1016/S0960-9822(03)00153-2.
- Loffler, G. (2014). Probing intermediate stages of shape processing. *Journal of Vision*, *14*(10):1458, doi:10.1167/14.10.1458. [Abstract]
- Loffler, G. (2015). Probing intermediate stages of shape processing. *Journal of Vision*, *15*(7):1, 1–19, doi:10.1167/15.7.1. [PubMed] [Article]
- Mante, V., Bonin, V., & Carandini, M. (2008). Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. *Neuron*, *58*, 625–638, doi:10.1016/j.neuron.2008.03.011.
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*, 605–610.
- McGovern, D. P., Hancock, S., & Peirce, J. W. (2011). The timing of binding and segregation of two compound aftereffects. *Vision Research*, *51*, 1047–1057, doi:10.1016/j.visres.2011.02.017.
- McGovern, D. P., & Peirce, J. W. (2010). The spatial characteristics of plaid-form-selective mechanisms. *Vision Research*, *50*, 796–804, doi:10.1016/j.visres.2010.01.018.
- Meese, T. S., & Freeman, T. C. (1995). Edge computation in human vision: Anisotropy in the combining of oriented filters. *Perception*, *24*, 603–622.
- Nam, J.-H., Solomon, J. A., Morgan, M. J., Wright, C. E., & Chubb, C. (2009). Coherent plaids are preattentively more than the sum of their parts. *Attention, Perception & Psychophysics*, *71*, 1469–1477, doi:10.3758/APP.71.7.1469.
- Pasupathy, A. (2014). Boundary curvature as a basis of shape encoding in macaque area V4. *Journal of Vision*, *14*(10):1456, doi:10.1167/14.10.1456. [Abstract]
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, *82*, 2490–2502.
- Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*, 1332–1338, doi:10.1038/nn972.
- Peirce, J. W. (2014). Compound feature detectors in mid-level vision. *Journal of Vision*, *14*(10):1455, doi:10.1167/14.10.1455. [Abstract]
- Peirce, J. W., & Taylor, L. J. (2006). Selective mechanisms for complex visual patterns revealed by adaptation. *Neuroscience*, *141*, 15–18, doi: S0306-4522(06)00563-X.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.
- Roach, N. W., & Webb, B. S. (2013). Adaptation to implied tilt: Extensive spatial extrapolation of orientation gradients. *Frontiers in Psychology*, *4*, 438, doi:10.3389/fpsyg.2013.00438.
- Roach, N. W., Webb, B. S., & McGraw, P. V. (2008). Adaptation to global structure induces spatially remote distortions of perceived orientation. *Journal of Vision*, *8*(3):31, 1–12, doi:10.1167/8.3.31. [PubMed] [Article]
- Robinson, A., & MacLeod, D. (2011). The McCollough effect with plaids and gratings: Evidence for a plaid-selective visual mechanism. *Journal of Vision*, *11*(1):26, 1–9, doi:10.1167/11.1.26. [PubMed] [Article]
- Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Washington, DC: Spartan.
- Schein, S. J., Marrocco, R. T., & de Monasterio, F. M. (1982). Is there a high concentration of color-selective cells in area V4 of monkey visual cortex? *Journal of Neurophysiology*, *47*, 193–213.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, USA*, *104*, 6424–6429, doi:10.1073/pnas.0700622104.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield. (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- von der Heydt, R., Peterhans, E., & Baumgartner, G.

- (1984). Illusory contours and cortical neuron responses. *Science*, *224*, 1260–1262.
- Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates, Inc.
- Wilson, H. R. (2014). From orientations to objects: Configural processing in the ventral stream. *Journal of Vision*, *14*(10):1460, doi:10.1167/14.10.1460. [Abstract]
- Wilson, H. R., & Wilkinson, F. (2015). From orientations to objects: Configural processing in the ventral stream. *Journal of Vision*, *15*(7):4, 1–10, doi: 10.1167/15.7.4. [PubMed] [Article]
- Wilson, H. R., Wilkinson, F., & Asaad, W. (1997). Concentric orientation summation in human form vision. *Vision Research*, *37*, 2325–2330.
- Yau, J. M., Pasupathy, A., Brincat, S. L., & Connor, C. E. (2013). Curvature processing dynamics in macaque area V4. *Cerebral Cortex*, *23*, 198–209, doi:10.1093/cercor/bhs004.
- Zeki, S. M. (1973). Colour coding in rhesus monkey prestriate cortex. *Brain Research*, *53*, 422–427, doi: 10.1016/0006-8993(73)90227-8.
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience*, *20*, 6594–6611.