Long-range interactions in visual perception

Lothar Spillmann and John S. Werner

Classical receptive-field concepts have been used to explain local perceptual effects such as border contrast and Mach bands, but are not sufficient to explain global perceptual effects. Examples are the perception of illusory contours, area contrast, color constancy, depth planes, coherent motion and texture contrast. These diverse effects require neurophysiological mechanisms within the visual pathways with long-range interactions. Candidate mechanisms are suggested, including converging feedforward projection to account for the emergence of new response properties at higher levels, recruitment of lateral connections to compensate for loss of afference and explain filling-in, and re-entrant projections from higher levels using synchronization of neuronal responses to account for binding.


IN THEIR PIONEERING STUDIES OF contrast phenomena, Mach1 and Hering2 in the last century proposed that each region of the retina might interact with many other distant regions (‘Wechselwirkung der Sehfeldstellen’). Fifty years later, the Gestaltists, in their attempt to understand phi motion (that is, apparent motion between two successive stimuli presented in different locations) and figure-ground segregation3, called for similar interactions in the realm of form and motion perception. In many demonstrations they showed that the appearance of a given part of the stimulus pattern depends on the context within which it is embedded. It is only recently that the anatomical and physiological mechanisms for these figure-ground relations (Wertheimer’s ‘Querfunktionen’ or lateral functions) have come under scrutiny.

The classical receptive-field concept originating with the work of Kuffler4 and Barlow5 provided the basis for contemporary ideas about neural representations of visual form. Yet, correlations between perception and center-surround antagonism of retinal receptive fields have been limited to short-range effects such as Mach bands6 and illusory spots in the Hermann grid (for further examples, see Refs 8-10). Such short-range interactions cannot explain more complex aspects of perception involving spatially separated areas in the domains of illusory contours, brightness, color, depth, motion and texture. Long-range interactions, primarily in the cortex, are required.

It is now accepted that in many visual neurons, stimuli presented outside the classical receptive field exert a strong and selective influence on the response to stimuli presented within the classical receptive field11. Among the earliest examples of long-range interactions was the periphery effect12, which showed that the response of retinal ganglion cells in the cat can be facilitated by moving a stimulus in the far periphery of the visual field. A similar interaction was demonstrated in retina and lateral geniculate nucleus with a shifting grating surround13,14. The perceptual significance of these effects is unclear, although they might play a role in visual attention and suppression during saccadic eye movements15. The problems discussed here, however, are not with physiological mechanisms in search of perceptual correlates, but with perceptual phenomena in quest of neuronal mechanisms.

In this paper, we present six phenomena requiring an explanation in terms of long-range interactions in the visual field. The mechanisms proposed to account for these interactions are: (1) converging feedforward projections leading to the generation of new receptive-field properties at higher levels; (2) horizontal intrinsic connections providing for interaction within a single visual area; and (3) re-entrant feedback projections diverging to cells at lower levels which respond selectively to specific properties of a stimulus. It has been argued that the divergent feedback associated with this latter mechanism might utilize synchronization of neuronal responses as the basis for binding. The six phenomena, all of which can be observed under natural conditions, and their hypothetical explanations by long-range interactions constitute a challenge for neurophysiologists as they shift their attention increasingly from local to global mechanisms of visual perception.

Illusory contours and brightness enhancement

In nature we rarely experience an unobstructed view of an object; most objects are seen only partially, due to occlusion by other objects. Despite incomplete views, however, there is seldom confusion about which parts belong together. A partially occluded object is perceived as unitary according to the Gestalt factors of good continuation, closure and the principle of ‘Prägnanz’ (figural goodness). Similar principles apply to the perception of an illusory figure in a seemingly incomplete stimulus configuration such as the Kanizsa triangle (Fig. 1), where an area of uniformly enhanced brightness bounded by illusory contours is generated by the corners of a hypothetical triangle. There is no luminance difference between the inside and outside of the perceived figure, yet the triangle appears to be lighter than the background. Although it is tempting to assign illusory contours (and surfaces) a role in everyday perception, they are in fact rarely if ever seen in a natural setting. Also,
Unlike recognition of a partially hidden object which will occur even if the occlusion is 'jagged', illusory contours require smooth gaps and closely aligned segments. Neurons with classical receptive fields cannot account for such contours; long-range connections within a larger 'response field' able to mediate perceptual completion across gaps of several degrees of visual angle are needed.

Peterhans, von der Heydt and Baumgartner\textsuperscript{16,17} found that approximately one third of the magnocells in area V2 of the macaque monkey, the second largest cortical area for visual processing, are responsive to cues eliciting the perception of subjective contours in human observers. End-stopped receptive fields of these cells are hypothesized to fill-in the gap and restore the (incomplete) Gestalt by connecting the parts with an illusory border. The parallel behavior between these cells and the percept suggests that one is a neural substrate for the other. Related phenomena implying long-distance interactions might be similarly explained. For example, illusory stripes ('phantoms') can be observed moving through an empty horizontal gap in phase with two vertical gratings on top and on the bottom\textsuperscript{18}. The uniform brightness enhancement observed in the Kanizsa triangle is an example of filling-in as discussed below.

Area contrast and filling-in of brightness and color

The fundamental cues defining figure–ground segregation in a stationary stimulus configuration are brightness, color, binocular disparity, motion and texture. In this section, we concentrate on the first two of these cues. Areas possessing the same brightness and color are usually seen as one, figure or ground. This segregation is supported by contrast mechanisms that enhance the difference between adjacent regions of the visual field. However, border contrast alone is not sufficient to describe our visual experience under normal conditions; enhancement of surfaces by area contrast is equally important. For example, a grey square of given luminance presented on different shades of grey will be perceived as having greatly different lightnesses. Or, a yellow disk if surrounded by an achromatic annulus of variable luminance will turn brown and ultimately black as the luminance of the surround is progressively increased above the luminance of the center ('Schwarzerhüllung'\textsuperscript{23,19}). A striking example of area contrast, known as colored shadows, was described by Goethe\textsuperscript{20}. His illustration is presented in Fig. 2. Here, an area of broad-band illumination, that would appear white under ordinary viewing conditions, looks bluish when surrounded by yellow. Unlike border contrast, which enhances the difference between figure and ground, area contrast serves to homogenize regions of different color and brightness within the figure.

What might be the processes and mechanisms responsible for brightness and color perception within uniformly stimulated areas bounded by an edge? Most neurons in area V1 and V2 do not respond to constant illumination within their receptive fields. Absence of response is ambiguous – it occurs in the presence of many stimulus conditions. Yet we do not see outlines, we see uniform surfaces bordered by contours. The need for signals from contours to prevent a perceptual void becomes apparent when the image is stabilized on the retina\textsuperscript{21}. Within seconds the entire percept will
disappear or fade (Troxler-effect; local adaptation). If, however, the stimulus contours are displaced, as with involuntary eye movements, the percept reappears. We conclude that in order to sustain the brightness and color of the central area, the signals corresponding to the edges must be constantly revived.

That these interactions depend on global, not local, mechanisms was shown by Gerrits and Vendrik, who observed that colors in stabilized retinal images encompassing the blind spot were perceived as uniform disks. This is a process known as filling-in. The scotoma corresponding to the physiological blind spot can be made visible only with special techniques, such as using small perimetric targets. If a scotoma is mimicked by a large uniform disk presented on a colored or textured background it will also tend to fill-in by acquiring the spatio-temporal properties of the surround. For example, a large red disk shown on a green background will rapidly disappear from vision to be replaced by a uniformly green surface if the border is blurred or stabilized on the retina. Similarly, a uniformly grey target on a dynamic noise background will temporarily fade from view to be filled-in by noise.

Recent evidence shows a neurophysiological correlate of this adaptation in area V3 (and to a lesser degree area V2) of the behaving monkey: within seconds of being presented with a textured stimulus containing an 'artificial scotoma' (produced by a white square), the initially lowered activity of a cell will recover and approach the activity elicited by the same stimulus, but no scotoma. The deficit of neural activity within the scotomatus area (and its presumed compensation by lateral interactions) is in a way analogous to filling-in of the blind spot. To explain these percepts, long-range signals arising at the edge and propagated from the boundary to the silent neurons representing the enclosed region are required. The ambiguity associated with the absence of neuronal response makes it possible for lateral interactions to impart a signal to those neurons that is consistent with the stimulus as encoded at the edge.

Color constancy

Color constancy refers to our ability to perceive a given color as the same despite changes in ambient illumination. This ability is essential for learning the identity of objects. It has long been known that the brightness and color of an area depend crucially on the illumination of the surround, particularly at the edges as was discussed for area contrast. In complex scenes, overall increases in illumination might be discounted by reciprocal lateral connections: the effects of light in the center are offset by the effects of light in the surround such that the net effect is to promote brightness and color constancy. The importance of integrating information over large regions of space to maintain color constancy was emphasized by Land, who showed that spatially complex stimulus patterns (patchworks of color called 'Mondrians' after the works of the Dutch painter Piet Mondrian) tend to maintain their appearance despite large changes in the intensity and spectral distribution of the light source. Shown in isolation, individual regions of the stimulus pattern would assume a different hue with changes in the intensity or spectral distribution of the illumination. Land proposed an algorithm for computing colors that would support constancy. This algorithm is based on a comparison of three receptor-specific lightness values from each surface in the scene with those of other surfaces. One way of implementing this algorithm would be to move a photosensitive 'mouse' along arbitrary trajectories from the perimeter of the Mondrian to the surface under consideration. By crossing a number of borders, a comparison can be made between the reflectance ratios for a given surface relative to that of neighboring surfaces. Alternative schemes have been proposed (for example, Ref. 30), but most retain the important notion that constancy depends on long-distance comparisons between different regions within a given pattern.

Long-range computations involved in color induction and filling-in of large areas might be generated already at the level of the retina or lateral geniculate nucleus, but the global computations required for color constancy most likely involve cortical interactions between both hemispheres. The importance of transcortical connections in color constancy was demonstrated in a patient whose corpus callosum had been transected surgically to control his epileptic seizures. For this individual, color constancy for stimuli falling on different hemifields was severely impaired relative to normal observers. This case study suggests that, given the appropriate stimuli, color constancy can be mediated by communication between the two hemispheres. Where might these long-range computations be carried out in the visual system? The answer is not known, but it is likely that they depend, at least in part, on area V4, where cell responses are modified by both local and global chromatic surround information including information coming from deep inside the ipsilateral visual hemifield. In primate area V4, Zeki has described 'perceptive' neurons that respond in a context-dependent manner, and seemingly consistent with the perceived color of a stimulus, rather than its wavelength. The inputs from the complex surrounds of these cells have not been quantified, but Zeki notes that they are long-range and defy description in terms of the classical receptive field.

The perception of depth planes

A phenomenon analogous to the uniform filling-in of bounded surfaces is the uniform induction of depth from the edges onto an enclosed area. This is called 'depth capture'. However, rather than by luminance or color contrast, the perception of depth planes is defined by the lateral disparities at the borders. For example, an outline square in which the vertical lines for one eye are displaced laterally (so as to produce crossed or uncrossed disparity) is perceived as uniformly elevated or recessed in depth in spite of the fact that the area between the lines provides no disparity information. It appears that the entire enclosed uniform surface assumes a different depth plane based only on the cues provided by the edges. This 'depth capture' is conceptually similar to the filling-in processes discussed earlier. To explain these effects, one has to assume long-distance interactions.

This assumption is also supported by the occurrence of depth contours in random-dot stereograms. Here, none of the monocular images contains contours to define a figure. The figure emerges only after binocular fusion has taken place (form-from-depth).
To extract depth from such a stimulus would seem to require a global comparison of the two monocular images. While there is evidence suggesting that the correspondence problem might already be solved by V1 neurons within local regions of the random-dot stereogram, the contour separating the depth planes appears to be made explicit only by cells in area V2 (Ref. 39). This contour almost certainly requires a global mechanism as it does not connect individual dots to make an irregular or jagged contour, but instead follows a perfectly smooth straight or curved line. Little is known about the algorithm used by the brain to perform this smoothing operation.

Coherent motion

Grouping of stimuli by coherent motion is a powerful effect to be attributed to binding. Figure 3 serves to illustrate how an invisible figure composed of dots on a dotted background will ‘pop out’ at the slightest movement by virtue of the Gestalt factor of common fate. The figure will also be perceived, but not as well, when only the background dots move or when both the dots representing the figure and those representing the ground move in different directions or at different speeds. Despite the coarseness and irregular distribution of the dots the word MOTION emerging from the noise field tends to have crisp, well-defined edges. It also is seen as lying slightly above the background. This is similar to the stratified sheets of noise sliding past each other when two groups of random dots moving in different directions are spatially superimposed. The high spatial resolution needed for the edge extrapolation and the low resolution needed for the stratification require mechanisms analogous to those that produce smoothing and filling-in of depth planes in random-dot stereograms. Single-cell studies in the macaque monkey have yielded neuronal correlates of motion segregation in area V1 (Ref. 41) and of motion transparency in the medial temporal area (area MT). Motion filling-in also occurs when one observes a large figure such as a homogeneous white square moving over a black background. In this example, motion signals originate only at the boundary of the figure while no signals are generated by the interior. Yet, we see the whole figure moving instead of just accretion at the leading edge and decretion at the trailing edge. This might be called motion capture by analogy to depth capture.

Using both psychophysical and electrophysiologically methods, Niedeggen and Wist showed that only a small percentage of spatially distributed, coherently moving dots in a dynamic random-dot field were sufficient to support figure-ground segregation. That coherent motion does not require identical directions and speeds for all elements in an array is demonstrated by biological motion such as observed when small lights are attached to the main joints of the body of a dancer in the dark. One can easily group the varying trajectories of the individual lights to perceive correctly the movements of the dancer. What might be the cellular basis of motion coherence? Britten and colleagues identified cells in area MT of the monkey that respond in a manner that is directly parallel to the behavioral findings. These cells have large receptive fields and are highly sensitive to the coherent translation of dots within a random-dot kinematogram. The ability of neurons from area MT and the medial superior temporal (MST) area to integrate motion information (including constriction, expansion and rotation) across large portions of the visual field requires substantial convergence of inputs from earlier cortical levels, such as V1 and V2. The generation of receptive fields that can mediate this long-range integration is likely to be accomplished through the convergence of feedforward connections.

Converging feedforward processing is also likely to mediate the perception of apparent motion as described by Wertheimer. Thresholds for apparent motion in humans and monkeys were measured by Newsome and colleagues, and their psychophysical results were compared with cell recordings in areas V1 and MT. The authors conclude that MT neurons are responsible for the perception of fast apparent movement, whereas both V1 and MT neurons are responsible for slow apparent movement. They suggest that the neuronal substrate for apparent motion might be distributed over multiple cortical areas, dependent on the speed and the spatial interval of the stimulus.

Texture contrast

There are several striking examples of texture segregation that suggest the involvement of long-range neuronal interactions across the visual field. The sudden emergence or ‘pop-out’ of a stimulus element embedded within a matrix of elements differing in orientation, color, polarity, size, shape and texture requires fast long-distance comparisons without any effort or conscious attention by the observer. These processes are independent of the number of distractors and have been attributed to parallel (versus serial) search. They are, therefore, called preattentive. It has recently been shown that neurons in area V1 respond depending on the perceptual context within
which the stimulus is placed. For example, the responses of many neurons were suppressed markedly by line textures extending beyond the classical receptive field\(^{51-53}\). These regions are normally silent, but will inhibit or facilitate the neuronal response if stimulated in conjunction with the classical receptive field. Accordingly, the response to an optimally oriented bar was enhanced if the surrounding texture was oriented orthogonally. In other words, responses of these cells were consistent with an increased perceptual salience, analogously to ‘pop-out’. Pop-out by orientation contrast is paralleled by a similar enhancement due to motion contrast when a cell’s response is facilitated by antiphase movement in its receptive-field surround\(^{44,45}\). Such neurons might be instrumental in mediating figure-ground segregation.

**Potential mechanisms**

Neurophysiological and anatomical studies have provided evidence that the visual system processes global information in at least three different ways: (1) converging feedforward projections from lower to higher levels; (2) recruitment of lateral connections within the cortex to compensate for loss of afference in deafferented cells; and (3) synchronization of neuronal responses and ‘tagging’ of lower-level cell activity by re-entrant signals. All of these mechanisms are relevant to the understanding of the phenomena described here.

Traditionally, hierarchical feedforward processing has been assumed whereby new receptive-field properties evolve from convergent inputs from lower-level neurons\(^{46-58}\). The idea here is that widely separated neurons at lower levels become interacting neighbors at higher levels through remapping. We now know that the projection from areas V1 and V2 to area MT and the inferotemporal cortex (IT) is not strictly topographic; spatially, the receptive fields of these areas are quite large with a substantial overlap. Barlow has argued that these non-topographic projections are not random, but occur in an orderly pattern according to stimulus dimensions (see also Ref. 59). He assumed that all information on one aspect of the stimulus (for example, motion in one direction, one velocity) is brought together from a wide region of the visual field. The large number of specialized pathways and cortical areas with their anatomically identified patterns of connectivity known to serve vision\(^{60,61}\) would support this bottom-up approach.

Among the six phenomena described, the filling-in of illusory contours and area brightness lends itself most clearly to an explanation in terms of converging feedforward projections. Peterhans, von der Heydt and Baumgartner\(^{62,63}\) propose that multiplicative inputs from end-stopped cells in area V1, tuned orthogonally, are required for generating the illusory contours of the Kanizsa triangle. A subjective contour is signaled when a higher-order cell in area V2 is activated concurrently by two or more of these distant inputs. Since the same neuron might also respond to a physical line in the same location as the subjective contour, the output of the cell is similar for the two types of stimuli, as is their appearance. Furthermore, when the sectors inducing a subjective contour are closed by thin curves, as illustrated in Fig. 1 (right), the neuronal response breaks down and so does the illusion. In the context of the model by Peterhans and colleagues, this would be expected because the closing lines fall on the inhibitory end-zones of the end-stopped cells and thus reduce their activation.

A second kind of neural processing to explain filling-in involves horizontal intrinsic connections within a single visual area. It is now known that dramatic dynamic changes occur in the size and location of cortical receptive fields when retinal activity is prevented from reaching its target area by deafferentation. For example, when a scotoma is induced in cat retinae by photocoagulation, receptive fields of cells in the lateral geniculate nucleus\(^{66}\) and striate cortex\(^{61,62}\) that had previously received input from the lesioned area become displaced and enlarged. This is illustrated in Fig. 4. The finding that the initially silenced neuron can be reactivated, within minutes, by stimuli falling onto the edge of the lesion\(^{64,66}\) suggests a mechanism that mediates input through lateral connections. An alternative explanation of short-term plasticity in terms of an overall increase in cell responsiveness, rather than a dynamic alteration of receptive-field structure, has been proposed\(^{67}\). The functional potentiation resulting in either case would lead to a perceptual filling-in of a gap, such as a retinal scotoma, although the much slower recovery in patients suggests that sprouting must also take place.

If deafferentation following a lesion leads to filling-in by lateral interaction, it would also be expected to occur when there is no patterned input due to a mask covering a certain region of the retina (artificial scotoma). Indeed, in the cat it has been shown that when the area from which a given neuron can be driven is masked by a black disk and a grating pattern is moved in the immediate surround, the receptive field expands up to five times its original size, enabling cells to become responsive again\(^{64}\). Similarly, following partial occlusion of a dynamic noise stimulus, cells in area V3 of the monkey will compensate rapidly for the reduced retinal stimulation with an increase in response, presumably due to an unmasking (disinhibition) of previously ineffective excitation from well outside the classical receptive field\(^{66}\).

An even larger expansion has been reported by Fiorani and colleagues\(^{69}\), who obtained responses in area V1 of the monkey when stimulating two regions
on either side of the optic disk that were 15 degrees apart. These distances are up to 10 times the lengths of conventional receptive fields, implying a functional (and structural) convergence much larger than hitherto thought. Although the blind spot is present from birth and thus represents a special case of a scotoma, the physiological findings suggest that our unawareness of it might be based on a filling-in process that uses mechanisms similar to the intracortical connections underlying filling-in of other kinds of scotomas.

Filling-in by horizontal interaction might also be responsible for uniform brightness and color perception in extended areas. One possible manifestation of this type of neural processing is illustrated in Fig. 5 for the Craik–O’Brien–Cornsweet illusion. Contrary to the luminance profile, the perceived brightness within the central area is uniformly low, whereas in the surround it is uniformly high. Clearly, these two brightness levels must be defined by the relative neural activities at the border. Horizontal connections of up to 6 mm between spatially separated hypercolumns have been found \(^{66,71}\) that could mediate propagation of edge signals and thereby assign different brightnesses and colors to regions of the visual field several degrees away. Polysynaptic cascades of local connections could carry these signals even further. It is important to add that the distribution of neural activity measured by microelectrode recording is only about 5% of that found by optical recording, in agreement with what might be needed to explain many long-range perceptual effects \(^{78}\). A similar interpretation applies to the segregation into smoothly delineated depth planes. Models assuming connectionist networks \(^{71}\) and boundary-triggered diffusion \(^{74}\) have been proposed to account for these effects.

A third neural process that supports global integration involves long-range feedback projections from higher areas to group cell responses at lower levels or across hemispheres \(^{75}\) by virtue of synchronization of neural activity among separate areas. These top-down projections need not be confined to the cortex. It is known, for example, that the feline cortex has massive projections to the lateral geniculate nucleus \(^{80}\). It should also be noted that top-down feedback from higher areas does not necessarily require synchronization, and conversely synchronization does not necessarily involve feedback from higher areas, but might be accomplished by feedforward or lateral interactions in some yet unknown manner.

While binding by re-entrant signals might contribute to more than just one of the phenomena described, it is most easily understood with respect to coherent motion. We know that spike activities between remote brain areas can interact by way of synchronization within a cell assembly, thereby producing a possible neuronal basis for perceptual coherence and connectivity \(^{77,78}\). Such a scheme would serve to enhance the salience of a figure with a minimum of neural connections: the significance of a neuron’s activity would depend on the context of activity in other neurons \(^{79,80}\). For the multitude of random dots comprising the word MOTION in Fig. 3, the synchronization of activity in cells with separate receptive fields might provide the binding necessary to obtain structure-from-motion. To confirm this hypothesis an experiment combining single-cell recording and behavioral testing is needed to be able to correlate the extent of synchronization with the perceptual salience of a stimulus. This experiment remains to be done.

What is the evidence for such a proposal? On average, each cortical area in monkey possesses connections to 10 others \(^{81}\). These connections are between areas at the same level as well as above and below in the cortical hierarchy, giving cells in any tissue patch distributed parallel access to information different in both retinotopic location and sensory specialization. Successful spike transmission to exploit these pathways requires synchronous discharge in widely separated areas. Striate and extrastriate neurons of both cat \(^{82}\) and monkey \(^{83}\) stimulated by coherent motion show synchronized discharges (short bursts occurring at 40–90 Hz), even though the individual stimulus elements are spaced over areas much larger than Hubel and Wiesel’s hypercolumns. This theoretical framework might also account for the long-range facilitation underlying the Gestalt factor of good continuation \(^{84,85}\).

Global integration might sometimes be based on more than one neural mechanism to produce a given phenomenon. For example, in a neural-network model, texture contrast has been attributed to feedforward processing with inhibitory interneurons between...
orientation detectors as well as to lateral modulation by orientationally tuned cells via horizontal connections. However, the same phenomenon could also be interpreted in terms of binding through synchronization. Similarly, while coherent motion and motion contrast were explained by synchronization, it might equally be accounted for by convergent feed-forward projections.

To explain many of the phenomena of the Gestalt school that have eluded neuroscientists for more than 50 years, it appears that we must deal with the possibility that simple percepts are the achievement of distributed processing by more than one cortical area. This idea requires two tasks from neuroscience: to establish mechanisms for the synchronization of information and binding among multiple areas, and to give reciprocal projections a more central role in spike transmission. The real challenge, however, lies in the problem of demonstrating that perceptual salience parallels the pattern of cell responses within a localized area, but, equally importantly, to the context in which they appear. This is consistent with the Gestalt credo that the whole is different from the sum of its parts.

Selected references


Acknowledgements

We are grateful for the valuable comments by H.B. Barlow, R. Eckhorn, A.K. Engel, C.D. Gilbert, S. Lehar, J.H.R. Maunsell, J.I. Nelson, P.H. Schiller and W. Sperling. JSW is a recipient of a Senior Scientist Award from the Alexander-von-Humboldt Foundation.