Beyond the classical receptive field: The effect of contextual stimuli

Lothar Spillmann
Department of Psychology, The University of Hong Kong, Pok Fu Lam, Hong Kong
Neurology Clinic, University of Freiburg, Freiburg im Breisgau, Germany

Birgitta Dresp-Langley
University of Strasbourg, ICube UMR 7357 CNRS, Strasbourg, France

Chia-huei Tseng
Department of Psychology, The University of Hong Kong, Pok Fu Lam, Hong Kong

Following the pioneering studies of the receptive field (RF), the RF concept gained further significance for visual perception by the discovery of input effects from beyond the classical RF. These studies demonstrated that neuronal responses could be modulated by stimuli outside their RFs, consistent with the perception of induced brightness, color, orientation, and motion. Lesion scotomata are similarly modulated perceptually from the surround by RFs that have migrated from the interior to the outer edge of the scotoma and in this way provide filling-in of the void. Large RFs are advantageous to this task. In higher visual areas, such as the middle temporal and inferotemporal lobe, RFs increase in size and lose most of their retinotopic organization while encoding increasingly complex features. Whereas lower-level RFs mediate perceptual filling-in, contour integration, and figure–ground segregation, RFs at higher levels serve the perception of grouping by common fate, biological motion, and other biologically relevant stimuli, such as faces. Studies in alert monkeys while freely viewing natural scenes showed that classical and nonclassical RFs cooperate in forming representations of the visual world. Today, our understanding of the mechanisms underlying the RF is undergoing a quantum leap. What had started out as a hierarchical feed-forward concept for simple stimuli, such as spots, lines, and bars, now refers to mechanisms involving ascending, descending, and lateral signal flow. By extension of the bottom-up paradigm, RFs are nowadays understood as adaptive processors, enabling the predictive coding of complex scenes. Top-down effects guiding attention and tuned to task-relevant information complement the bottom-up analysis.


doi: 10.1167/15.9.7  Received February 15, 2015; published July XX, 2015
ISSN 1534-7362 © 2015 ARVO
perception of extended areas segregated from their surroundings. Together they account for phenomena such as the perception of uniform surfaces, filling-in, and grouping (Grossberg & Mingolla, 1985; Spillmann & Werner, 1996; Lamme, Super, & Spekreijse, 1998). These mechanisms have been proposed to enable the transition from local to global processing by using information from beyond the classical RF. Surfaces enclosed by boundaries, rather than edge effects, have become the main focus of interest. A cellular basis for these mechanisms will have to be sought in cortical visual areas rather than in interactive processes in the retina (Spillmann, 1997, 1999). Detailed laminar cortical models of figure–ground segregation have been proposed to unify the explanation of many perceptual and neurobiological data about such boundary–surface interactions (e.g., Cao & Grossberg, 2005).

Long-range interaction between RFs serves not only unperturbed everyday perception, it also provides for “repair” mechanisms when sensory information is incomplete or ambiguous, such as in the perceptual filling-in of scotomata (Spillmann & De Weerd, 2003; Spillmann, 2011), the completion across the physiological blind spot (Kawabata, 1984; Komatsu, 2006, 2011), and in predictive scene coding of natural images.

Beyond the classical RF

The earliest attempts to relate visual physiology to perception were severely limited by the techniques available. Wolfgang Köhler (Köhler & Held, 1949), one of the fathers of Gestalt psychology, together with Held, invoked “electric field” effects to explain how the perception of patterns would be produced in the brain. In their study, they set out to demonstrate an isomorphic shape correlate of pattern vision (see the review by Wurtz, 2009), but what was missing at the time was an appropriate technology for recording interpretable brain signals that could lead to an understanding of the neuronal mechanisms underlying perception. Yet, Köhler and Held’s (1949) concluding remarks in that paper, suggesting that access to the cortical correlates of complex pattern vision would have an immediate impact on any theory of psychophysics and perception, turned out to be prophetic.

Hardly a decade later, single-cell recordings from the cat and monkey brains (Hubel & Wiesel, 1959, 1962, 1965, 1968) were to produce exactly such an impact, marking the beginnings of a deeper understanding of the ways in which information is passed on from one processing stage to another in the brain. Systematic investigation of RF properties did not stop there. Input effects from outside the classical RF were soon discovered (for review, see Allman, Miezin, & McGuinness, 1985b), leading to a distinction between local and global percepts; the definition of contextual stimuli; and, in the 1990s, the concept of an association field (Field, Hayes, & Hess, 1993). This latter concept was a psychophysical blueprint for linking up stimulus elements lying on a common path. Therewith, Köhler and Held’s (1949) idea that shape perception could find an explanation in terms of a global brain field theory was taken to the next level. This review deals with this higher level by recounting how the study of long-range signal interaction between cortical neurons has produced theoretical developments beyond the classical RF with new concepts for understanding the neural basis of complex scene integration in the brain.

Long-range interaction and contextual neurons

Our perception relies on the interaction between proximal and distant points in visual space, requiring short- and long-range neural connections among neurons responding to different regions within the retinotopic map. Evidently, the classical center-surround RF can only accommodate short-range interactions; for long-range interactions, more powerful mechanisms are needed. Accordingly, the hitherto established local RF properties had to be extended to take distant global inputs into account.

The idea of an extended (called nonclassical or extraclassical today) RF was not new. Kuffler (1953, p. 45) already wrote, “… not only the areas from which responses can actually be set up by retinal illumination may be included in a definition of the receptive field but also all areas which show a functional connection, by an inhibitory or excitatory effect on a ganglion cell. This may well involve areas which are somewhat remote from a ganglion cell and by themselves do not set up discharges.” The first evidence for a distant modulation of a neuron came from McIlwain (1964), who demonstrated in the cat that a moving stimulus in the far periphery of the RF enhanced the response to a stimulus localized within the RF, i.e., the periphery effect. Next Blakemore, Carpenter, and Georgeson (1970) and Blackmore and Tobin (1972) in the cat showed that lines of different orientation interacted antagonistically, suggesting mutual inhibition between neighboring cortical columns. In a follow-up paper, Rose and Blakemore (1974) targeted a specific inhibitory neurotransmitter (bicuculline) to account for this effect. Thereafter, Fischer and Krüger (1974) in the lateral geniculate nucleus (LGN) demonstrated that a grating jerk in the far surround of an RF produced a brisk neuronal excitation in the center, i.e., the shift effect. This discovery was followed by reports in the cat...
cortex of an unresponsive or silent surround (Maffei & Fiorentini, 1976) and, more importantly, a region beyond the classical RF, generating interactive effects between coaxial lines (Nelson & Frost, 1978). Yet, von der Heydt, Peterhans, and Baumgartner (1984) were the first to find neurons in V2 of the monkey cortex that responded to an ”incomplete” bar as if receiving input from outside its classical RF. The authors interpreted this response as a mechanism designed to bridge a gap in a discontinuous contour (see also Peterhans, von der Heydt, & Baumgartner, 1986).

Following these early discoveries, researchers started using contextual stimuli to study context-sensitive neurons (Allman, Miezin, & McGuinness, 1985a; Gilbert & Wiesel, 1990; Gilbert, 1992; Knierim & Van Essen, 1992; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995). Figure 2 illustrates two examples. On the top (left) is shown a pattern with a small vertical bar embedded within a textural background of horizontal bars, i.e., orientation contrast. A neuron in cat area 17 responded much more strongly to this pattern than to the uniform control pattern on the right, in which all bars have the same orientation (Kastner, Nothdurft, & Pigarev, 1999). Evidently, cross-orientation between the central bar and the bars in the surround enhanced the response whereas iso-orientation inhibited it. The same relationship is obtained for the pattern shown on the left (bottom), in which the center bar moved in one direction while the bars in the surround moved in the opposite direction, i.e., motion contrast. Again, the neuronal response to this pattern was much stronger than the response to the control pattern on the right, in which all bars moved in the same direction. In both examples, the difference in relative rather than absolute response level enables the pop-out (Li, 1999, 2002). Note that the stimulus surround for both kinds of patterns in Figure 2 exceeded the size of the classical RF. Neurons therefore must have received input from beyond this area (see also Jones, Grieve, Wang, & Sillito, 2001; Jones, Wang, & Sillito, 2002).

To illustrate the various ways in which RF surrounds influence their centers, we here present some of the most compelling examples of contextual modulation from the ever-growing literature.

In a psychophysical experiment on contour integration in human observers, Field et al. (1993) tested the detectability of a string of Gabor patches (Figure 3, left) aligned on a background of randomly oriented Gabor patches (Figure 3, right). The authors varied (a) the angle of element rotation relative to the path, (b) the angle of path deviation from collinearity, and (c) the distance of neighboring Gabor patches from each other. Deviation from collinearity affected detectability the most, suggesting that the Gestalt factor of good
continuation was critical for contour integration. Remarkably, the string could still be detected when the distance between the aligned elements was five times the length of an individual Gabor patch. Field et al. interpreted their findings in terms of long-range interaction within an association field. Their data confirmed the prediction of how bipole receptive fields may complete boundaries (Grossberg, 1984; Grossberg & Mingolla, 1985).

Importantly, trained rhesus monkeys produced psychophysical thresholds similar to those of human observers (Mandon & Kreiter, 2005). Such long-range interaction likely involves contextual RFs at different retinal locations, which are grouped to-

Figure 2. Contextual patterns for orientation contrast (top left) and motion contrast (bottom left). In the experiment, bright bars were used on a dark background. The center bar was located inside the classical RF (small rectangle) of a neuron in cat area 17, and the surround bars were positioned outside the RF. Mean responses for motion contrast are also shown (right). The response to the contrast patterns (columns 2 and 6) was in the same order of magnitude as that to the center presented in isolation (columns 0 and 4), and there was hardly any response to the surround shown by itself (columns 3 and 7) and no response either to the uniform patterns (columns 1 and 5) serving as a control. (Modified from Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1999). Neuronal responses to orientation and motion contrast in cat striate cortex. Visual Neuroscience, 16, 587–600.)

Figure 3. Perceptual grouping of aligned elements according to the Gestalt factor of good continuation. A string composed of collinear Gabor elements is easily perceived when shown in isolation (left) but is hard to detect when embedded in a background of randomly oriented Gabor patches (right). In the experiment, the “snake” on the left had not previously been shown to the observer. (From 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.)
together by higher-order neurons in the extrastriate cortex.

Another experiment demonstrates the influence of contextual modulation by using collinear facilitation. Figure 4 (top) shows that the response of a V1 neuron to a low-contrast test line in the RF was enhanced when a high-contrast collinear flanker was presented outside the RF. The flanker itself (middle) did not elicit a response. Bottom: Psychophysical threshold of a human observer in the absence (thin curve) and presence of the flanker (thick curve). The leftward shift of the response curve (arrow) indicates facilitation. (Modified from Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15, 843–856.)

Consistent with their neuronal results, Kapadia et al. (1995) also showed in human observers that the psychophysical threshold for raising a subliminal line to visibility was lowered by the presence of a collinear flanker, i.e., facilitation (Figure 4, bottom). Stimulus conditions were critical: A small lateral offset of the two lines from collinearity, a difference in relative orientation, or a short orthogonal line between the target line and the flanker weakened and ultimately abolished the facilitating effect. These findings are compelling evidence that contextual modulation works similarly at both the single neuron level and the population level, proving Köhler’s early “field” intuitions right.

Several authors reported comparable effects (e.g., Dresp & Bonnet, 1991; Dresp, 1993; Polat & Sagi, 1993, 1994; Yu & Levi, 1997, 2000; Chen, Kasamatsu, Polat, & Norcia, 2001; Tzvetanov & Dresp, 2002; Dresp & Langley, 2005; Huang, Chen, & Tyler, 2012). Some of the effects shown therein were found to depend on the contrast intensity of the stimuli with facilitating interactions at low flanker contrast and inhibitory interactions at higher flanker contrast intensities (e.g., Polat & Norcia, 1996; Wehrhahn & Dresp, 1998; Chen & Tyler, 2001, 2008).

Horizontal interactions in area V1 are known to be of shorter range than in V2, and although the neurophysiological data summarized in Figure 4 show that such interactions can enhance neuronal responses to short aligned stimuli, long-range boundary completion as demonstrated for illusory contour formation in monkey (von der Heydt et al., 1984; Figure 1) is limited predominantly to neuronal processing in area V2 (V1 in cat, Redies, Crook, & Creutzfeldt, 1986). A model by Grossberg, Mingolla, and Ross (1997), simulating the von der Heydt et al. (1984) and Kapadia et al. (1995) data, illustrates this distinction between V1 and V2.

The third experiment exhibiting contextual modulation from the surround is based on neuronal processing of orientation contrast as a prerequisite for figure–ground segregation in the monkey (Lamme, 1995). Figure 5 (left) shows the stimulus display: two test patches of line segments with opposite or same orientation to the background. While the patch on the right is barely discernible, the one on the left merges with the background and is invisible. The same stimuli are illustrated schematically by iso- or crosshatched windows (located within the four boxes in the middle). The RF (small black rectangle) of a V1 neuron was always fully enclosed within the test patch. Thus, the neuron should not have received any input to inform it of the orientation of the surrounding background. The results indicated otherwise: Lamme found that the neuron readily discriminated between cross-orientation and iso-orientation of the test patch. The response to cross-orientation (graphs on the right) was always
stronger. This finding implies that the first steps of figure–ground segregation may already be built into the responses of the earliest, retinotopically mapped, cortical area. The figure–ground enhancement effect, however, occurred with a delay of 30–40 ms, suggesting that feedback from higher visual areas may play a role in this mechanism.

A last experiment to be mentioned here involves a phenomenon called border ownership. According to Rubin (1915/1921), a figure occludes the ground and “owns” the border surrounding it. In an experiment tapping the neuronal mechanism of border ownership, Zhou, Friedman, and von der Heydt (2000) found neurons predominantly in V2 (but also V1) of the monkey that responded selectively to the location of a figure relative to the RF. For example, an edge-selective neuron responded strongly to a contour when it was owned by a figure on the neuron’s preferred side and significantly less to a contour that was owned by a figure on the other side (Figure 6). Note that the contrast step in the RF is the same for all six configurations shown, but the direction in terms of figure–ground is opposite in Figure 6A and B. The neuron illustrated has a preference for border ownership to the lower left, but other neurons with the same RF location showed the opposite preference. This suggests that any contour is represented by two groups of neurons with opposite border ownership preferences. These studies are among the clearest demonstrations of contextual influences from beyond the classical RF. By varying the distance of the remote contours from the RF (e.g., by varying the size of the squares), it is possible to measure the extent of the contextual influence (Zhang & von der Heydt, 2010, their figure 5).

Figure 5. Figure–ground segregation in area V1 of the trained, alert monkey. (Left) Stimulus display. Only the patch on the right-hand side is visible due to orientation contrast to the ground, and the patch on the left merges with the background. (Middle) Schematic representation of the RF (small black rectangle) within a 4° × 4° window whose hatching is either cross- or iso-oriented to that of the background. Black and white demarcations were not shown. Fixation was on the center dot. Note that the orientation of the hatching in (a) and (b) is the same as is the hatching in (c) and (d). (Right) Neuronal responses were significantly larger when the hatching within the window was cross-oriented to the background than when it was iso-oriented. (Modified from Lamme, V. A. (1995). The neurophysiology of figureground segregation in primary visual cortex. The Journal of Neuroscience, 15, 1605–1615.)
Border ownership selectivity and side preference are intrinsic properties of the individual neuron, possibly based on modulatory feedback from hypothetical “grouping cells” at a higher level (Craft, Schütze, Niebur, & von der Heydt, 2007; Mihalas, Dong, von der Heydt, & Niebur, 2011). A recent study finding elevated spike synchrony between border ownership neurons when activated by the same object, which provides strong evidence for such feedback (Martin & von der Heydt, 2015). Selective attention to a figure was found to enhance the responses representing border ownership (Qiu, Sugihara, & von der Heydt, 2007).

These examples show that neuronal responses depend not only on local stimulus analysis within the classical RF, but rather on global feature integration, and these contextual influences can extend over relatively large regions of the visual field (Gilbert & Li, 2013). This adds further proof to the Gestalt credo that a whole is not reducible to the sum of its parts. Likewise, the response of a cell to a complex stimulus cannot be fully predicted from the responses to its elements (Gilbert, 1992). Suddenly, the RF was recognized as fulfilling an important role for figure–ground segregation and surface perception, which are critical steps of processing for object perception and recognition (Nakayama, He, & Shimojo, 1995).

Contextual influences in vision and visual perception have attracted increasing interest in psychophysical and neurophysiological research (Albright & Stoner, 2002; Series, Lorenceau, & Frégnac, 2003; Li & Chen, 2011), prompting the question: How large is the outer field of such contextual neurons? Zipser, Lamme, and Schiller (1996), using contextual modulation for a textural figure, found that the spatial extent of contextual modulation of a parafoveal RF was approximately 8°–10° of visual angle.

Measurements by Bringuier, Chavane, Glaeser, and Frégnac (1999) based on intracellular recordings reported similar orders of magnitude. These authors demonstrated that the visually evoked synaptic integration field in cat cortex extends over an area four to 15 times the size of the classical RFs of Hubel and Wiesel (1962). An even larger figure comes from a study of Angelucci, Levitt, and Lund (2002a), who suggest on anatomical grounds that the field of contextual influence is 20 times larger than the classical RF. Figure 7 illustrates the classical RF center, the (classical) near surround, and the (extraclassical) outer surround. The authors attribute the first to feed-forward from the LGN, the second to horizontal input from within V1 (cortico–cortical), and the third to feedback from extrastriate areas (Hupé et al., 1998; Angelucci et al., 2002a, 2002b).

RF size varies not only by virtue of contextual interaction with the outer surround; it also varies with retinal eccentricity (Daniel & Whitteridge, 1961; Drasdo, 1977) and area within the visual hierarchy. Smith, Singh, Williams, and Greenlee (2001) have compiled average data from the literature on single-unit recordings in the monkey, showing that classical RFs increase in size from near foveal to peripheral locations (for computational modeling, see Schwartz, 1980) and from V1 to higher areas in the extrastriate cortex (Gattass, Sousa, & Gross, 1988; Freeman & Simoncelli, 2011). RFs are smallest in the primary visual cortex (V1), larger in V2, larger again in V3/VP, and largest of all in areas V3A and V4 (Figure 8A). Also the slope of the functions describing the increase in size with eccentricity increases progressively from lower to higher visual areas. Dumoulin and Wandell (2008) likewise present average data for neurophysiological RFs (single- and multiunit) from the literature and
arrive at the same stacking order although the slopes of their regression lines for visual areas V1, V2, and V3 are less steep (Figure 8B).

For processing visual information, RFs are not isolated entities of single neurons, but part of an interrelated network, in which one input affects another. Functional magnetic resonance imaging (fMRI) has recently been used to map neuronal responses to complex patterns and estimate the population receptive field (pRF) in various visual field locations. These quantitative estimates of pRF size in occipital regions of the human visual cortex are shown in Figures 8C through F. Overall data for V1–V3 compare reasonably well with single-cell RF measurements obtained at corresponding eccentricities and locations in monkey visual cortex (Figure 8A, B), while RF sizes in LO1 (lateral occipital) are much greater. Ordinates and abscissas in Figure 8 have been scaled appropriately for better comparison except for Figure 8C, the ordinate of which is given in duty cycles (percentage).

Compared with RFs of neurons in V1–V3, RFs in yet higher visual areas, such as the inferotemporal (IT) and middle temporal (MT) lobes are considerably larger, covering as much as 25° of visual angle (Felleman & Kaas, 1984; Rolls, Aggelopoulos, & Zheng, 2003); they also lose much of their retinotopic organization although this has been disputed for human brains (Wandell & Winawer, 2011). At the same time, such neurons encode increasingly complex stimuli. For example, although the RFs of neurons in areas V1–V3 mediate perceptual filling-in, contour integration, and figure–ground segregation, neurons in IT respond to faces (Perrett, Rolls, & Caan, 1982) and in MT to coherently moving patterns (Desimone, Albright, Gross, & Bruce, 1984; Singer, 1989) and biological motion (Oram & Perrett, 1994). For a summary of visual percepts and their presumed level of origin in the brain, see table 1 in Spillmann (2009).

Furthermore, RF properties of cells in lower cortical areas are rather fixed compared to those in the temporal and parietal cortex, which are more malleable (Ben Hamed, Duhamel, Bremmer, & Graf, 2002; Quraishi, Heider, & Siegel, 2007).

### Dynamic RF topography: Changes in RF size and location

Ever since Hartline’s (1938, p. 410) first description, RFs of single cells were assumed to be invariant in size and location. Although this is generally true, it does not hold for RFs of cells that are deprived of their input. Gilbert (1992; see also Gilbert & Wiesel, 1992) reported that in the cat following a focal retinal lesion, RFs of cortical neurons fell silent immediately after deafferentation as was expected. However, within minutes, these same neurons responded again when light fell on the regions next to the lesion. At the same time, RFs near the lesion boundary expanded (by a factor of up to five) and shifted outward from the lesion site, implying dynamic changes in both RF size and location. This is shown in Figure 9. The change in cortical topography of RFs suggests that neurons can be quickly “rewired,” presumably by recruiting collaterals through disinhibition. Long-range interaction would then enable them to respond to input from outside the lesion area for which they were previously unresponsive. Gilbert and Wiesel (1992) in the monkey (Lund et al., 1993) and cat have reported long axonal connections capable of propagating information from the edge to the interior of a given area.

The remapping of RFs from positions inside the lesion area to locations partly outside has been proposed as a possible mechanism for perceptual filling-in across a scotoma (Spillmann & Werner, 1996). In normal vision, horizontal interactions of this kind might also underlie the induction of brightness and color contrast, assimilation (neon color, watercolor effect) and their relationship to perceived stratification and transparency (Gilbert, 1992). For example, there is evidence that neurons stimulated by the edge of a surface actively propagate their information to neurons...
Figure 8. Comparison of RF and pRF measurements by various authors. Sizes are plotted as a function of retinal eccentricity for visual cortices V1–V4. Top row (A): Single cell RFs in the monkey, average data from the literature. (B) Estimates derived from single- and multiunit activity and local field potentials in nonhuman primates. Solid lines indicate averages from the literature. Middle and bottom rows (C–F): pRFs derived from fMRI measurements in human subjects. Sources: (A) Smith et al., 2001; (B) Dumoulin & Spillmann, Dresp-Langley, & Tseng (2015) 15(9):7, 1–23.
representing the interior of that surface via long-range interaction. In this way, filling-in from the border may sustain the brightness of the enclosed surface area (Huang & Paradiso, 2008). Also, when a steady stimulus of uniform luminance was shown within the classical RF while the background luminance was modulated well beyond the RF area, the response of the neuron to the uniform stimulus could be modified (Rossi, Rittenhouse, & Paradiso, 1996). DeValois, Webster, DeValois, and Lingelbach (1986) studied this effect psychophysically (see also Paradiso & Hahn, 1996). Spatial interactions between distant patches of retina reinforce Mach (1865) and Hering’s (1878) assertion that knowing what is locally activated is not sufficient for predicting what is globally perceived (Spillmann, 1997).

Gilbert and Wiesel (1992) found that several months after the lesion, the RFs had migrated even further to the outside of the lesion scotoma. This lesion-induced shift in location suggests that the cortex of the adult cat possesses considerable plasticity and is capable of “repairing” a hole (scotoma) in the visual field although at the cost of geometrical topography. Figure 10 illustrates the resulting distortion. Spillmann and Werner (1996) suggested that the perception of brightness, color, texture, and stereo depth in and across a scotoma might conceivably be restored from the surround by virtue of such a mechanism. Interpolation of image features from the surround may also account for perceptual completion across the physiological blind spot. Fiorani, Rosa, Gattas, and Rocha-Miranda (1992) obtained responses in area V1 of the monkey when stimulating two regions on opposite sides of the optic disk that were 15° apart. These retinal distances are several times the spatial extent of conventional RFs, implying a functional (and structural) convergence much larger than hitherto thought. Yet, in terms of cortical magnification, the same distances might be greatly reduced. Komatsu (2011) has proposed a hypothetical wiring diagram, which accounts for spreading information in the blind spot region by intracortical circuitry (Figure 11). According to this account, retinal signals from the region surrounding the BS are conveyed from layers 2/3 of V1 to layer 6, where they form a large RF (solid circle), providing completion and filling-in of the void. Whether this diagram also accounts for oriented filling-in (Kawabata, 1984) remains to be shown.

Cortical models (Grossberg, 1994, 1997; Cao & Grossberg, 2005; Grossberg & Yazdanbakhsh, 2005) exploiting the functional properties of laminar cortical organization, first demonstrated by the pioneering work of Ramón y Cajal (1899), have provided physiologically inspired mechanistic models to account for the completion and filling-in of gaps across stimuli. Therein the authors distinguish between two distinct, but complementary, mechanisms: (a) perceptual completion of boundaries, which is orientation-selective, and (b) surface filling-in, which is unoriented. A good example of this fundamental distinction is the Kanizsa triangle, which is perceived both by virtue of its amodal contours and the brightness enhancement of the enclosed surface.

Similar mechanisms of neuronal activity may also apply to the filling-in of an artificial scotoma, i.e., a uniform surface in the absence of any lesion. Pettet and Gilbert (1992) recorded dynamic changes in cat RF size...
when they covered the RF of a cortical neuron with a uniform mask on a jittering background. In analogy to the retinal lesion condition, the neuron responded to stimulation from outside the mask with a fivefold increase in size. This is illustrated by Figure 12, in which the RF size originally corresponded to frame size #1 but expanded to frame size #2 when an occluder covered the RF. When the occluder was removed, the RF shrank to frame size #3, just to re-expand to frame size #4 when it was put back on again. Finally, without the occluder, the RF collapsed to frame size #5, slightly smaller than its original extent (frame #1). Such changes in size occurred within a span of only 15 min after applying the mask, suggesting unmasking of preexisting connections.

These results show that RF size adjusts itself to stimulus demands and, thereby, challenging two established beliefs in neuroscience: (a) that the RF would correspond to an invariant set of photoreceptors funneling their inputs onto higher-level sensory neurons and (b) that there would be a fixed RF map based on retinal topography.

Figure 10. Hypothetical “repair” of a retinal scotoma caused by retinal laser coagulation. (a) Visual hemifield with hole representing the scotoma. (b) Cortical representation of scotoma. (c) Scotoma in V1 is gradually closed at the cost of retinal and visual field topography. (From Gilbert, C. D., & Wiesel, T. (1992). Receptive field dynamics in adult primary visual cortex. *Nature, 356*, 150–152.)

Figure 11. Hypothetical wiring diagram for the perceptual filling in of the blind spot. (A) Cortical representation of the blind spot of the right eye when both eyes are open. The blind spot, clearly visible by the empty notch at the level of the LGN, is completely obscured in V1 due to input from the left eye. (B) Cortical representation of the same area, when the left eye is closed. Here the area corresponding to the blind spot is “filled in” at the level of V1 by feedback from neighboring layers 2/3 onto layer 6. In both cases, the blind spot is not perceived. (From Komatsu, H. (2011). Bridging gaps at V1: Neural responses for filling-in and completion at the blind spot. *Chinese Journal of Psychology, 53*, 413–420.)
Dilks, Baker, Liu, and Kanwisher (2009) recently reported similar results in psychophysics. These authors “deprived” the region of the visual cortex (V1) corresponding to the blind spot in one eye by patching an observer’s contralateral eye. Within seconds of this deprivation, observers reported a white square 0.5° away from the boundary of the blind spot to extend (“stretch”) into the blind spot. This perceptual elongation is suggestive of rapid RF expansion within the deprived blind spot area in V1, analogous to findings from single-cell recordings after a retinal lesion (Gilbert & Wiesel, 1992). A similar effect was observed in a patient who had suffered a stroke that destroyed the fibers that normally provide input to the upper left visual field in V1 and who described a black square presented to the lower left visual field as a “finger” reaching toward and into the upper blind visual field (Dilks, Serences, Rosenau, Yantis, & McCloskey, 2007). Similarly, a circle was described as cigar-like and a triangle as pencil-like.

These fleeting perceptual elongations were confirmed by fMRI (Dilks et al., 2007) and are consistent with the assumption of a cortical reorganization in V1 due to long-standing deafferentation. In this context, it is noteworthy that Ricco’s area for complete spatial summation (the psychophysical equivalent of a neuronal RF) becomes larger by 30% when retinal cell density decreases with age (Malania et al., 2011). On the other hand, no evidence for cortical remapping has been found in patients suffering from age-dependent macular degeneration (AMD) (Baseler et al., 2011). This may be because, here, lateral separation in cortical tissue space is exceptionally large due to the foveal magnification factor. On the other hand, filling-in of a line, grating, or regular dot pattern in AMD patients has been found to occur across several degrees of visual angle (Zur & Ullman, 2003), suggesting long-range cortical reorganization in V1.

Since the early discoveries of extraclassical RF effects, their functional characteristics have been studied further to explore how bottom-up mechanisms, such as end-stopping, would account for the long-range coding of feature discontinuities in visual stimuli. Findings revealed that the firing rates of cat cortical neurons in area 17 exposed to edges perpendicular to their preferred orientation were enhanced when a “feature border” was presented outside and close to the RF (Li 1999, 2002; Schmid, 2008). The “feature borders” were defined by discontinuities in phase, orientation, or motion direction of the stimulus. A comparison with control measures led to the conclusion that the enhanced firing rates were due to a release in suppression (i.e., disinhibition). Model accounts of the observations suggest that center–surround interactions, contextual modulation, and end-stopping are part of a single brain mechanism for representing spatial discontinuities in visual image analysis, by virtue of which, quite often, several goals must be achieved simultaneously as in orientation-based texture segmentation (Schmid & Victor, 2014).
Although visual RFs are typically considered bottom-up detectors, or neuronal filters, selective only to given stimulus parameters (Spillmann, 2014), contextual neurons have recently been found in area V4 that are modifiable by attention, i.e., top-down processing (Krause & Pack, 2014). Specifically, the allocation of spatial attention may be understood as a behavioral characteristic of visual RFs (Treue & Martinez-Trujillo, 2012), whose sensitivity to spatial stimuli is dynamically modulated by the attentional spotlight. There are two modes of attention: passive and active. For example, a perceptual object in the visual field may capture attention in a stimulus-driven fashion, or it may become subject to goal-directed top-down attentional control (e.g., Yantis & Jonides, 1990, Conci, Töllner, Leszczynski, & Müller, 2011).

Attention is one modifier of RFs; the choice of stimuli is another. Most studies cited so far used laboratory stimuli. With the advent of studies in alert monkeys using free viewing and natural stimuli, an increase in information transmission efficiency has been found for natural scenes in V1 (Gallant, Connor, & Van Essen, 1998; Vinje & Gallant, 2002). These studies suggest that extraclassical RF effects may be linked to the predictive coding of natural images. Neural circuits would learn the statistical regularities of the natural world and communicate deviations from regularity to higher processing centers. Such selective signaling would reduce redundancy by discarding the predictable, hence redundant, components of the input signal (Rao & Ballard, 1999). More recently, interpretations of extraclassical RF effects have been extended even further in computational models, suggesting that V1 firing patterns may deliver universal signatures of visual saliency irrespective of their intrinsic feature preferences, e.g., contrast polarity (Li, 1999, 2002). What is more salient is likely to figure and to attract visual attention first.

Despite these advances into uncharted territory, much of the response variance in V1 still remains unexplained, and it is likely that one of the central functions of cortical processing is to predict upcoming stimulus events based on the spatial and temporal context of a scene. Muckli and colleagues (Muckli, Vetter, & Smith, 2011; Muckli & Petro, 2013) investigated the information content of feedback projections using the apparent motion path between two alternating stimulus locations (Wertheimer, 1912) or an illusory shape suggesting a partially occluded triangle (Kanizsa, 1955) to probe for a response in retinotopic regions of the brain. This is reminiscent of Ginsburg (1975), who used spatial filtering in an attempt to isolate and enhance the illusory triangle in a Kanizsa figure, thereby demonstrating, for the first time, that the relevant information (i.e., the illusory contours) generated by the incomplete stimulus pattern was implicit in the overall spatial relationships of that pattern (for review, see Dresp, 1997). Similarly, Muckli et al. (2011) and Muckli and Petro (2013) analyzed fMRI activity patterns corresponding to incomplete stimulus parts and found that they generated object-related percepts.

The authors concluded that extraclassical RFs of neurons in V1 carry information relevant to the conscious interpretation of an incomplete stimulus as a meaningful whole. This kind of predictive coding introduces a conceptual shift in visual neuroscience, with which the brain is seen as continually generating models of the world based on information from memory in order to give meaning to incomplete sensory input. As pointed out already by Helmholtz (1867/1924), MacKay (1956), and Gregory (1980), our perception is guided by inferences, or object hypotheses, by which it seeks to resolve ambiguities in the stimulus in the most plausible manner (Spillmann & Dresp, 1995). In the brain, such predictive models would be created in higher cortical areas and communicated to lower areas through feedback connections (Muckli & Petro, 2013).

### Outlook and perspectives

Research on RFs started 75 years ago and is moving on swiftly. During the last 20 years, the RF concept in neuroscience research has undergone a complete revision from that of the earlier years (for review, see Löfler, 2008; Spillmann, 2014), showing that functional properties of RFs depend (a) on the visual context in which a target stimulus is embedded and (b) on the method of analysis used. This article does not attempt to be complete in reviewing all the extraclassical RF effects reported in this rapidly evolving field. Rather, we restricted ourselves to describing some of the major findings in the literature. Developments clearly do not stop here, and further exciting discoveries will undoubtedly come up in the near future.

In a nutshell, although RFs were formerly believed to have invariant response characteristics, they are, in fact, modifiable by intracortical (lateral, recurrent) interactions (e.g., Bair, 2005; Yeh, Xing, Williams, & Shapley, 2009; Neri, 2011; see also Grossberg & Raizada, 2000). Also, Hubel and Wiesel's (1962, 1965) initial distinction between complex and hypercomplex cells in the functional hierarchy of the primary visual cortex had to be reconsidered (Mechler & Ringach, 2002; Bair, 2005). Simple stimuli, such as flashed spots, oriented bars, and drifting gratings, used in the early studies of RF properties only revealed the most basic response properties. Our knowledge of RF size and location in different parts of the visual field has since
evolved considerably (see Figures 7 and 8). Hartline’s (1938) dictum that RFs are “fixed” has been shown to be untrue. Furthermore, it was found that one and the same stimulus feature elicits a stronger response when embedded in a natural scene rather than in a random field (Field, 1987). Complex stimuli, such as natural images or movie clips (Olshausen & Field, 1996; Gallant et al., 1998; Vinje & Gallant, 2002; Felsen & Dan, 2005), have revealed new RF substructures (see also Ringach, Hawken, & Shapley, 2002; Carandini et al., 2005; Schwartz et al., 2012).

The recent proposal that V1 responses constitute visual saliency maps (e.g., Li, 1999, 2002; Zhaoping, 2008, 2014) adds to the early intuitions by Köhler and Held (1949) relative to the existence of a Gestalt field at the level of neural representation. The research on predictive coding (e.g., Muckli & Petro, 2013) discussed above has provided us with new accounts of the functional role of complex intracortical feedback and top-down processing. The RFs of what were formerly called feature detectors are influenced by spatiotemporal context, selective attention, and memory. Why this is so can be understood on the basis of the brain’s need to constantly update knowledge. Not only do familiar objects need to be detected and recognized quickly, new objects never before encountered need to be learned as well and added to the memory inventory. This requires that the already learned visual representations are stable and accessible and that there is capacity for the processing and acquisition of new and not yet learned input. Grossberg (1983) called this the stability–plasticity dilemma, which is addressed by his adaptive resonance theory (ART). In ART networks, top-down projections generate a hypothesis for the recognition of objects from the sensory input. If such a hypothesis is recurrently reinforced and consolidated, it is believed to lead to conscious recognition (Grossberg, 1983, 1997).

RFs thus have acquired an important role in providing knowledge about the visual world. The ecological relevance of RF behavior was first touched upon by Barlow (1953), Lettvin, Maturana, McCulloch, & Pitts (1959), Jung (1961), and Baumgartner (1990), and it is coming into the focus of contemporary neuroscience. The key questions here are (a) how do RFs change dynamically to enhance their contribution to visual perception in different tasks, and (b) how does the visual brain integrate local cues to form global representations within a dynamically changing world (see von der Heydt & Peterhans, 1989; Spillmann, 1999; Pan et al., 2012).

Thus, the revised RF concept takes into consideration not only functional plasticity and a bottom-up saliency map, but also top-down processes, such as spatial attention, the detection of irregularities, scene recognition, and priming. Such modulation by higher-level input becomes plausible if one considers that—surprisingly—far more fibers descend from the primary visual cortex (V1) of the monkey to the LGN than ascend in the opposite direction (Peters, Payne, & Budd, 1994). The results of a systematic study using localized tissue cooling (Payne, Lomber, Villa, & Bullier, 1996) are consistent with massive feedback from higher visual areas (V4, V3) to lower ones (V2, V1).

Spatial attention, for example, would appear to act like a gain control mechanism, enhancing the perceptual salience of the object under consideration and suppressing information from outside the focus of interest (Itti & Koch, 2001). Recent research (e.g., Li, 1999; Gilbert & Li, 2013; Schmid & Victor, 2014) suggests that, in addition to responding to select physical properties of local stimuli, RFs and their associated neurons avail themselves of information from global stimuli that are relevant to the ongoing perceptual task. Rather than possessing fixed functional properties, as suggested by Hartline (1938), RFs are therefore conceived as dynamic processors whose tuning changes according to stimulus context, expectation, and attention.

For example, in a behavioral curve-tracing task, it was demonstrated that neurons, whose RFs lay along a curved contour, responded more strongly when the contour was attended to by the monkey rather than when it was unattended (Roelfsema, Lamme, & Spekreijse, 1998). These data show that attention and perceptual grouping interact in the interest of boundary formation as predicted by laminar cortical models of vision (e.g., Grossberg & Raizada, 2000; Raizada & Grossberg, 2001). How cortical processes subserving boundary formation interact with the top-down processes that control attention is one of the core issues addressed by these models. Poort et al. (2012), recording from V1 and V4 in the monkey, conclude that boundary detection is an early process based on bottom-up computation whereas surface filling occurs later, requires feedback, and is facilitated by visual attention.

It thus appears that the RFs of extrastriate neurons behave like matched filters, or templates, that are dynamically tuned to optimize visual processing and visual search (David, Hayden, Mazer, & Gallant, 2008; Schmid & Victor, 2014). Their selectivity for searched patterns is enhanced by attention (Itti & Koch, 2001; Ipata, Gee, & Goldberg, 2012). It has long been known that a cell in the superficial layers of the superior colliculus responds more robustly when a stimulus that falls within its RF becomes the target for a subsequent saccade (Goldberg & Wurtz, 1972). We now know that attention, in conjunction with goal-directed saccades, modulates the RFs of neurons in macaque V4 and MT by shifting their centers toward attended locations, not unlike a flashlight (Colby, Duhamel, & Goldberg, 1996; Tolias et al., 2001; Womelsdorf, Anton-Erxleben, & Treue, 2008; Womelsdorf, Anton-Erxleben, Pieper, &
This spatiotemporal dynamics in predictive remapping demonstrate that, already in the extrastriate cortex, RF properties are not invariant but highly adaptable, enhancing perceptual processing whenever a task requires it.

The attentional spotlight tells the brain where in the restricted space of the visual field a change has occurred. This is called detection. But for recognition (is it a line, a dot, or a small animal?), top-down processes are needed to identify the perceived objects. Beyond the level of signal detection, perception relies in part on information stored in memory representations (e.g., Churchland, 2002). Future research on extraclassical RF properties will have to include studies on processes of perceptual learning and memory. The temporal firing characteristics of neurons are critical in these processes (e.g., Jensen, Idiart, & Lisman, 1996; Churchland, 2002) as most of perceptual learning is temporal rather than spatial (see Wang, Cong, & Yu, 2013).

Résumé

Since Hartline’s (1938, 1940) original studies in the frog, the RF concept has evolved in several ways. Table 1 includes the most important discoveries: (a) lateral inhibition; (b) simple, complex, and hypercomplex (end-stopped) RFs; (c) RFs sensitive to selective rearing and deprivation, critical period for neuronal plasticity; (d) difference of Gaussians (DOG) filters, Fourier channels; (e) perceptive fields; (f) RFs beyond the classical RF; (g) contextual fields; (h) RF dynamics; and (i) RFs responsive to natural scenes.

Hubel and Wiesel (e.g., Hubel, 1963) in the early 1960s advanced the hypothesis that RFs of cells at a given level of the visual system emerged by combining a number of lower-level RFs. Sustained and transient channels in the cat were added to their hierarchical feed-forward concept of simple, complex, and hypercomplex cells in the 1970s. The feed-forward concept was challenged in the mid-1980s, when researchers discovered that neuronal responses could be modified by stimulation from the extended outer surround (i.e., beyond the classical RF). In the 1990s, further research reported long-range horizontal interactions as well as reciprocal projections from higher visual areas, enabling higher level neurons to modulate neuronal responses at lower levels by feedback (Hupé et al., 1998; Cudeiro & Sillito, 2006).

Our understanding of visual perception has gained immeasurably from each of these steps. Population perceptive fields (pRFs) have become the psychophysical and perceptual counterparts of RFs. A next step is the application of the RF concept to natural stimuli and natural scenes, including cognitive strategies, such as attention, search, and perceptual learning (W. Li, Pièch, & Gilbert, 2004, 2008; Seitz & Dinse, 2007; Wang et al., 2013). It thus seems that after 75 years of research, bottom-up finally meets top-down and feature detection gives way to perception and cognition.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Major discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hartline, 1938, 1940</td>
<td>Summation area in frog optic nerve</td>
</tr>
<tr>
<td>Jung, von Baumgartner, &amp; Baumgartner, 1952</td>
<td>First single cell recording from cat cortex</td>
</tr>
<tr>
<td>Barlow, 1953</td>
<td>Lateral surround in frog optic nerve, fly detector</td>
</tr>
<tr>
<td>Kuffler, 1953</td>
<td>Lateral surround in cat optic nerve</td>
</tr>
<tr>
<td>Barlow et al., 1957</td>
<td>Loss of lateral inhibition in scotopic vision</td>
</tr>
<tr>
<td>Lettvin et al., 1959</td>
<td>Bug detector</td>
</tr>
<tr>
<td>Jung, 1961</td>
<td>Psychophysical correlates of visual neuron responses</td>
</tr>
<tr>
<td>Hubel &amp; Wiesel, 1962</td>
<td>Simple cells</td>
</tr>
<tr>
<td>Hubel &amp; Wiesel, 1965, 1968</td>
<td>Complex cells, end-stopped cells, hypercolumn</td>
</tr>
<tr>
<td>Blakemore &amp; Cooper, 1970</td>
<td>Effects of selective rearing and stimulus deprivation</td>
</tr>
<tr>
<td>Enroth-Cugell &amp; Robson, 1966; Campbell &amp; Robson, 1968</td>
<td>DOG, Fourier channels, contrast sensitivity function</td>
</tr>
<tr>
<td>Jung &amp; Spillmann, 1970</td>
<td>Perceptive fields and psychophysical correlates</td>
</tr>
<tr>
<td>Ransom-Hogg &amp; Spillmann, 1980; Oehler, 1985</td>
<td>Westheimer function in man and monkeys</td>
</tr>
<tr>
<td>Nelson &amp; Frost, 1978</td>
<td>Responses from beyond the classical RF</td>
</tr>
<tr>
<td>Baumgartner et al., 1984</td>
<td>Perceptual neurons for filling-in illusory contours</td>
</tr>
<tr>
<td>Knierem &amp; Van Essen, 1992</td>
<td>Contextual stimuli</td>
</tr>
<tr>
<td>Field et al., 1993</td>
<td>Contextual neurons</td>
</tr>
<tr>
<td>Kapadia et al., 1995</td>
<td></td>
</tr>
<tr>
<td>Lamme, 1995</td>
<td></td>
</tr>
<tr>
<td>Sillito et al., 1995</td>
<td></td>
</tr>
<tr>
<td>Kastner et al., 1999</td>
<td></td>
</tr>
<tr>
<td>Gilbert &amp; Wiesel, 1992; Pettet &amp; Gilbert, 1992</td>
<td>Filling-in of lesion scotoma and artificial scotoma</td>
</tr>
<tr>
<td>Gilbert &amp; Li, 2013</td>
<td>RFs as dynamic processors</td>
</tr>
<tr>
<td>Angelucci et al, 2002a, b Olshausen &amp; Field, 1996</td>
<td>Large outer surround</td>
</tr>
<tr>
<td></td>
<td>RFs and natural stimuli</td>
</tr>
</tbody>
</table>

Table 1. Select steps in RF research. Authors are given on the left, major findings on the right.
Acknowledgments

This work was supported by a General Research Fund from the Research Grants Council of Hong Kong, China, and the Hong Kong University Seed Funding Program for Basic Research to Dr. Tseng. It represents an expanded version of an invited talk given at APCV 2013 in Su-zhou, China. The talk was sponsored, in part, by a Teaching Excellence Fellowship of the University of Hong Kong (to Prof. Chia-huei Tseng), research funds from the Peking-Tsinghua Center for Life Sciences, Peking University (to Prof. Cong Yu), and Chinese Academy of Sciences, Shanghai (to Prof. Wei Wang). Profs. Horace Barlow, Charles Gilbert, Barbara Heider, Wu Li, Susana Martinez-Conde, Jerry Nelson, Leo Peichl, James Thomas, John Troy, Wei Wang, and John S. Werner, and Li Zhaoping offered valuable comments on earlier versions of the manuscript. Moeka Komachi, Jing Ting Huang, and Dr. Franz Aiple helped with the figures, Matt Oxner and Diederick Niehorster with editorial work. The librarians of the Medical Library at the University of Freiburg provided valuable assistance in bibliographical research. We thank them all.

Commercial relationships: none.
Corresponding author: Chia-huei Tseng.
E-mail: CH_Tseng@alumni.uci.edu.
Address: Department of Psychology, The University of Hong Kong, Pok Fu Lam, Hong Kong.

References


Mach, E. (1865).


Rubin, E. (1915/1921). *Synsoplevede Figurer*. Copenhagen: Glyndendalske [Translated as Visuell wahrgeommene Figuren (Copenhagen 1921)].


Schmid, A., & Victor, J. D. (2014). Possible functions of contextual modulations and receptive field


