

Perspective

Solving the binding problem: Assemblies form when neurons enhance their firing rate—they don't need to oscillate or synchronize

Pieter R. Roelfsema^{1,2,3,4,*}¹Department of Vision & Cognition, Netherlands Institute for Neuroscience (KNAW), 1105 BA Amsterdam, the Netherlands²Department of Integrative Neurophysiology, VU University, De Boelelaan 1085, 1081 HV Amsterdam, the Netherlands³Department of Psychiatry, Academic Medical Centre, Postbus 22660, 1100 DD Amsterdam, the Netherlands⁴Laboratory of Visual Brain Therapy, Sorbonne Université, INSERM, CNRS, Institut de la Vision, 17 rue Moreau, 75012 Paris, France*Correspondence: p.roelfsema@nin.knaw.nl<https://doi.org/10.1016/j.neuron.2023.03.016>

SUMMARY

When we look at an image, its features are represented in our visual system in a highly distributed manner, calling for a mechanism that binds them into coherent object representations. There have been different proposals for the neuronal mechanisms that can mediate binding. One hypothesis is that binding is achieved by oscillations that synchronize neurons representing features of the same perceptual object. This view allows separate communication channels between different brain areas. Another hypothesis is that binding of features that are represented in different brain regions occurs when the neurons in these areas that respond to the same object simultaneously enhance their firing rate, which would correspond to directing object-based attention to these features. This review summarizes evidence in favor of and against these two hypotheses, examining the neuronal correlates of binding and assessing the time course of perceptual grouping. I conclude that enhanced neuronal firing rates bind features into coherent object representations, whereas oscillations and synchrony are unrelated to binding.

INTRODUCTION

When we open our eyes, we immediately see what is there. The efficiency of our vision is a remarkable achievement of evolution. The introspective ease with which we perceive our visual surroundings masks the sophisticated machinery in our brain that supports visual perception. The image that we see is rapidly analyzed by a complex hierarchy of cortical and subcortical brain regions. Neurons in low-level brain regions extract basic features such as line orientation, depth, and the color of local image elements.¹ They send the information to several mid-level brain areas. For example, neurons in area MT code for motion direction,² and neurons in area V4 code for color and shape fragments.^{3–6} Neurons in mid-level areas send the information to yet higher levels for an even more abstract analysis of the visual scene.⁷ Neurons at these higher levels code for the category of objects and even for the identity of specific individuals.^{8–10} Hence, every visual object activates a complex representation that is carried by a large set of neurons across many brain regions.

An important question is how the distributed and fragmented representations of objects across many areas of the visual brain can lead to a unified perception of objects against a background (Figures 1A and 1B). I focus here on this so-called “binding problem.” The binding problem occurs if there are multiple objects. Each of the objects activates a pattern of neurons across many brain regions, and in such a representation it may not be evident which features belong to one of the objects and which

ones belong to the others. Which process glues the features into coherent object representations? Neuronal mechanisms for binding have been controversial. Initially, researchers proposed binding by synchrony (BBS), which holds that the binding problem is solved by synchronous oscillations.^{11,12} A related hypothesis, known as communication through coherence (CTC), holds that synchrony and oscillations permit separate channels of communication between different brain regions.^{13,14} I will review evidence that casts doubt on oscillations and synchrony as the code for binding and communication. Instead, the binding problem appears to be solved by a coordination of neuronal firing rate levels across brain regions. The representations of features of an object become linked when the neurons coding for them enhance their firing rates (binding by firing rate enhancement [BBRE]).^{15,16} There is a close correspondence between the firing rate enhancement and object-based attention, a concept from perceptual psychology.^{17,18} Later sections will discuss how the propagation of firing rate enhancements in the visual brain corresponds to the spread of object-based attention throughout an object's representation until all of its elements are attended to and thereby bound in perception.

I will first outline why there is a binding problem. I will then evaluate BBS and BBRE based on the experimental evidence in favor of and against these two binding mechanisms. Toward the end of the review, I will also briefly touch upon the role of feature binding for motor programming, cognitive routines, and its relation to workspace theories of conscious perception.



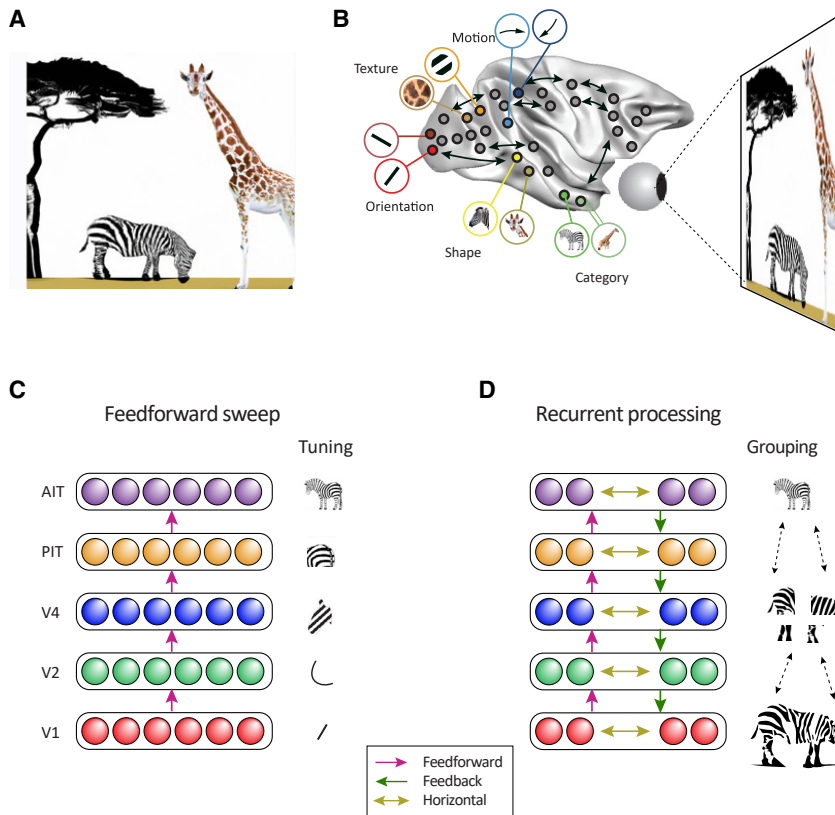


Figure 1. Two phases of visual processing
 (A) Image generated by OpenAI's image-generation network DALL·E2 using the prompt "a zebra and a giraffe and a tree, realistic."
 (B) Schematic indicating how visual features are represented in a distributed manner, with features like motion and texture in different brain regions. A macaque brain is illustrated, but distributed representations are present in many species. Neurons (colored circles) are positioned at approximate locations of feature representations.
 (C) Visual processing starts with a feedforward sweep during which low-level brain regions encode simple features and propagate the activity to higher regions, which extract more complex features up to the level of object categories.
 (D) In a later, recurrent response phase, feedback (green arrows) and horizontal connections (yellow arrows) come into play. Perceptual grouping operations link distributed representations into coherent object representations.

HARD-CODED FEATURE CONJUNCTIONS AND FLEXIBLE CODES FOR BINDING

The processing of a new image starts with a phase during which feedforward connections propagate information from lower to higher cortical areas so that increasingly abstract features are extracted.¹⁹ During the early phase of neuronal responses, information is propagated from lower to higher visual brain regions that encode increasingly complex features (Figure 1C). Deep neural network models²⁰ provide insight into how neuronal networks, including the visual system, can create meaning out of retinal inputs.^{21,22} Although many deep neural network studies focus on object categorization, other feature hierarchies exist in the brain, for example for the analysis of complex motion patterns and for the coordinate transformations that are required to control body movements during actions like navigation and grasping.^{7,23,24} If the task is object recognition, feedforward processing may suffice for simpler scenes, and perceptual grouping operations may not be required.^{25,26} However, there are many tasks in which it is essential to determine which of the lower- and higher-level features are part of the same object.

There are at least two solutions for this binding problem. The first is that groupings are hard coded during feedforward processing. Many neurons in visual cortex code for feature conjunctions, which have been called "base groupings."¹⁶ A neuron might respond, for example, only to red vertical lines, thereby coding that these features belong together.²⁷ Neurons

in higher brain areas coding for concepts, like zebras and giraffes, also code for feature conjunctions. A neuron that represents zebras, for example, codes for feature constellations that include a zebra head, legs, trunk, and stripes. Hence, such a neuron codes for a perceptual group, although it can be activated by many zebra pictures. However, note that the specific zebra of Figure 1A is unique, and it is unlikely that there is a neuron that responds only to this specific configuration of features. Establishing which low-level contours belong to the zebra requires a second form of perceptual grouping that is flexible and is called "incremental grouping." Furthermore, there is a limit to the total number of feature conjunctions that can be hardwired in the visual brain. There is a "combinatorial explosion"¹¹ as more feature conjunctions are possible than the number of neurons in the visual brain. Both BBS and BBRE propose that these additional conjunctions are formed by an incremental grouping process, which tags features of one object and segregates them from features of other objects. This incremental grouping process can occur for feature constellations that were never seen before and is thought to rely on feedback connections from higher to lower visual areas and horizontal connection between neurons in the same area (Figure 1D).^{16,19}

Incremental grouping works for unfamiliar shapes, which are not represented as base groupings and can hence only be represented in a distributed manner. For example, in Figure 2A, one can see that the two parts indicated by red arrows belong to the same object, whereas the green arrow points to a different one. Unfamiliar shapes are represented by an assembly of neurons responding to features that are linked by lower-level grouping cues, such as the collinearity of image elements, their connectedness, and similarity of color and texture. These low-level grouping cues were described by Gestalt psychologists²⁸⁻³⁰ and explain why we can see where one unfamiliar

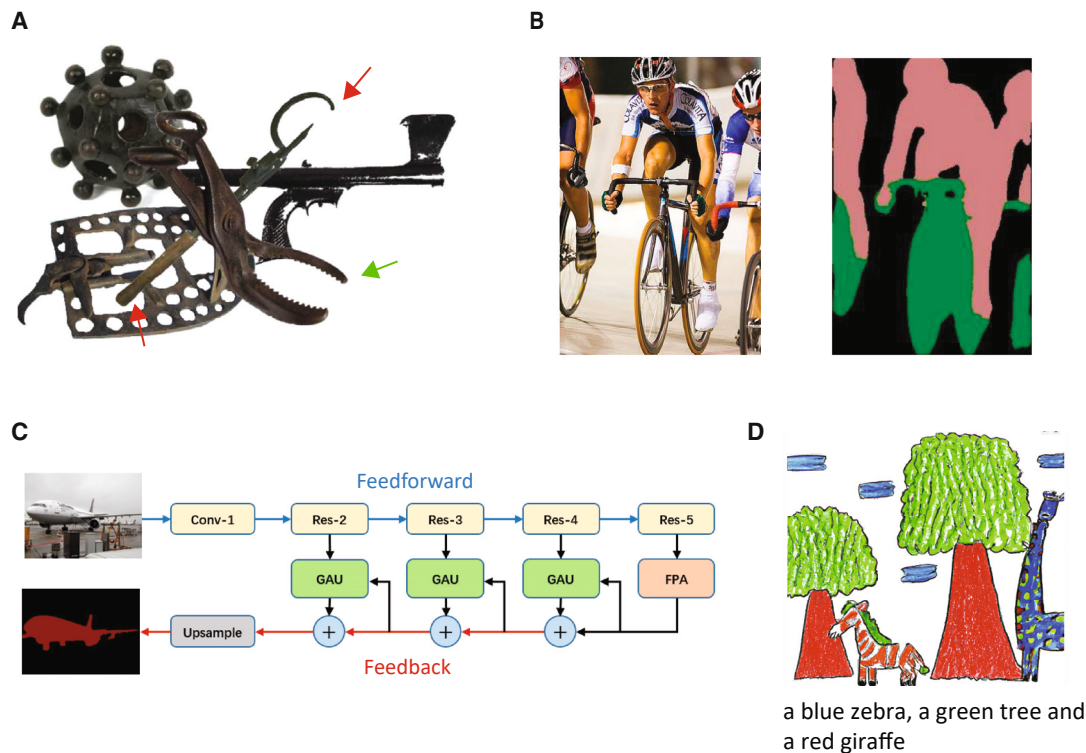


Figure 2. Perceptual grouping of unfamiliar objects and semantic segmentation

(A) Superimposition of unfamiliar objects. One can see that the two red arrows point to the same object, whereas the green arrow points to a different one. (B) Semantic segmentation labels pixels that belong to object categories, here, people and bicycles. (C) Example “pyramid attention network” for semantic segmentation. Processing starts with a feedforward pass (using a ResNet) where increasingly abstract features are extracted up to the level of object categories. This is followed by a feedback pass in which the features and pixels that belong to object classes are labeled differently. GAU, global attention upsample; FPA, feature pyramid attention. (D) Erroneous binding by DALL·E2. Cued with the prompt “a blue zebra, a green tree, and a red giraffe,” the network created a blue giraffe and a red zebra. (B) and (C) are reproduced from Li et al.³¹

object ends and another one starts. Incremental grouping is flexible because one can choose to focus on any of the objects and segregate it from the others.

In the machine vision field, the problem of image segmentation has received a lot of interest in recent years. Figure 2B illustrates the output of an example neural network that performs semantic segmentation. This means that it identifies pixels belonging to different object categories.³¹ The example network labeled the pixels that are part of humans with one color and the pixels of bicycles with another color. There are interesting similarities between semantic segmentation by neural networks and the perceptual grouping operations that take place in the human brain. The artificial neural network of Li et al.³¹ starts with a forward pass in which the units extract increasingly abstract feature constellations in higher layers up to semantic categories (Figure 2C). This is followed by a feedback pass that labels the pixels that belong to specific object categories. Whereas semantic segmentation gives the same label to pixels of all objects if there are multiple instances of a class, there exist related “instance segmentation” networks.³² These networks give different labels to separate objects of the same category and would segregate the three cyclists in Figure 2B as different objects. The deep neural networks for image segmentation may

provide crucial insights into the mechanisms for image parsing in human vision, and I will point out correspondences and differences in later sections.

Incremental grouping can go awry in both humans and neural networks. In humans, illusory conjunctions between features of different objects can occur if there is not enough time for recurrent processing. In a classic study, Treisman and Schmidt³³ asked participants to report two digits in a briefly presented display. A few colored letters were placed between the digits, and the subjects also reported the identity and color of the letters as a secondary task. The subjects sometimes reported a letter with the color of a different letter. For example, the subject might report a brown T, although a blue T and a brown R had been presented, suggesting that these features were correctly registered but not bound. Interestingly, such binding errors are also produced by neural networks, such as OpenAI’s deep network DALL·E2, which generates images from text. The network sometimes produced a red zebra and a blue giraffe when cued to produce the opposite combination of colors and shapes (Figure 2D). These illusory conjunctions illustrate the fundamental nature of the binding problem caused by the distributed and overlapping representation of multiple objects.

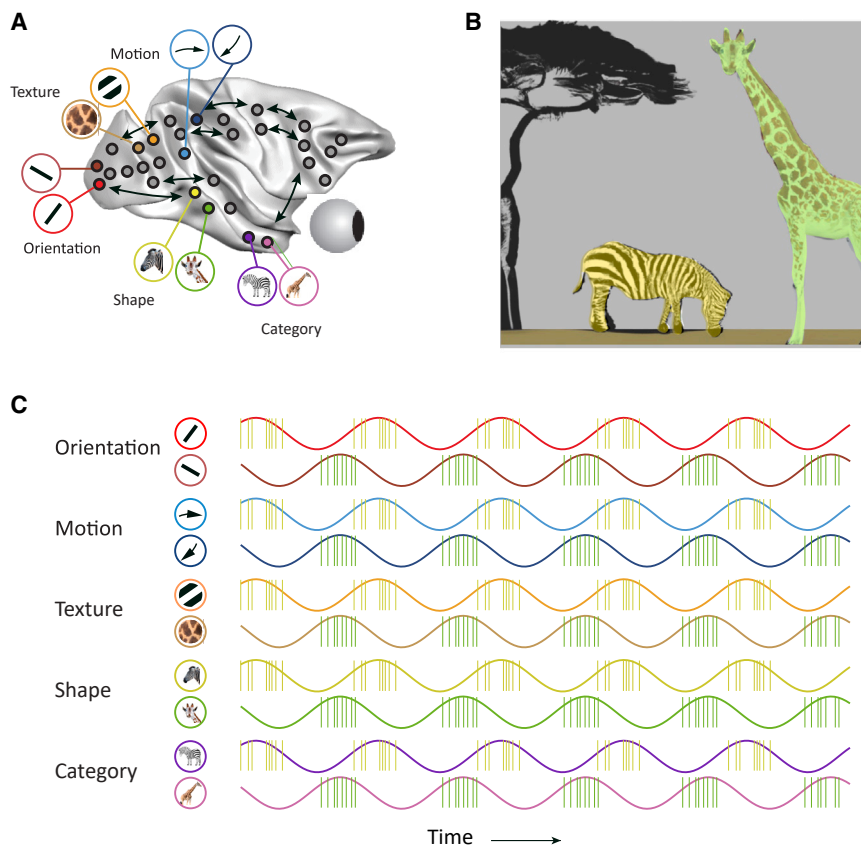


Figure 3. Oscillatory synchrony as a mechanism for binding

(A) Distributed representation of visual features. (B) BBS permits multiple incremental groups. Oscillatory synchrony labels features of the zebra and the giraffe, which are both segregated from the background (gray). (C) Neurons representing features of the same object fire action potentials in synchrony.

surfaces.³⁷ Fries et al.³⁸ measured synchrony in V1 in cats during binocular rivalry. In this paradigm, different stimuli are presented to the two eyes, and the stimulus that is perceived alternates between the eyes. Synchrony was stronger for V1 neurons coding for the stimulus that dominated perception, suggesting a link between gamma synchrony and conscious perception. Gamma oscillations also increase when stimuli are attended,^{39–41} although these increases are not always restricted to the gamma band.⁴² It has been argued, however, that these increases in gamma may be related to the firing rate increases that accompany attention shifts given the coupling between gamma power and neuronal firing rates.⁴³ The BBS hypothesis quickly gained popularity in neural network modeling studies,^{44,45} and the

OSCILLATIONS AND SYNCHRONY DO NOT BIND

More than thirty-five years ago, von der Malsburg and Schneider³⁴ proposed that neuronal synchronization binds features into coherent object representations. The BBS hypothesis holds that neurons coding for features of the same object fire in synchrony, whereas neurons coding for features of different objects do not.^{12,35} Neurons coding for features of one object (e.g., the zebra in Figure 3) would fire action potentials in synchrony and out of sync with spikes fired by neurons coding for other objects such as the giraffe (green vs. yellow spikes in Figure 3C). An important feature of BBS is that the representations of objects of interest carry their own temporal tag so that multiple incremental groups can coexist in perception (green and yellow in Figure 3B). An additional advantage of BBS is that synchrony provides a coding dimension in addition to the firing rate, enhancing the expressiveness of the representation.

A study in the primary visual cortex of cats by Gray et al.³⁶ provided initial evidence that gamma oscillations (30–80 Hz) might provide a rhythm for binding. Neurons in area V1 of cats exhibited stronger synchrony when they coded for features of a single, elongated bar of light than if they responded to two distinct and shorter bars. A related study in support of BBS recorded neuronal activity in areas 18 and postero-medial bank of the lateral suprasylvian sulcus (PMLS) of the cat and revealed that synchrony increased for neurons encoding features of the same surface compared with when they responded to different

ideas were also extrapolated to functions outside vision, such as reading.⁴⁶

However, there are now several findings that seem incompatible with BBS (for earlier criticisms, see Shadlen and Movshon and Reynolds and Desimone, respectively^{47,48}). First, gamma synchrony is a local phenomenon. Its strength quickly decays to zero for neurons separated by distances larger than a few mm.^{49–51} Gamma synchrony can therefore not group image elements that are represented by neurons with a larger separation. For example, neurons in human V1 representing the legs and head of the giraffe in Figure 3B can be separated by centimeters, depending on the gaze position and the size of the picture on the retina. Therefore, gamma oscillations cannot be used as a code for semantic segmentation. The same holds for the binding of features across brain regions. Interregional gamma synchronization between brain regions is relatively weak.⁵² To enhance sensitivity for the detection of synchrony, researchers often resort to analyzing local field potentials (i.e., the local electroencephalogram [EEG]) in one or both areas.^{41,53,54} Only a few studies analyzed the gamma synchronization between spikes recorded between different areas. One of these studies by Jia et al.⁵⁵ recorded spikes in area V1 and in layer IV of area V2, which receives the feedforward input from V1. Spikes in the areas only exhibited weak synchronization if the neurons had overlapping receptive fields. The spikes in V2 were not in synchrony but lagged those in V1 by approximately 3 ms, which is consistent with the feedforward propagation delay between V1 and layer

IV of V2.⁵⁶ If neurons in successive stages would exhibit similar delays, they fire out of phase with the V1 neurons, which is problematic for BBS. The weak interareal gamma synchronization is problematic for setting up independent communication channels between brain regions, as is proposed by CTC.^{13,14} Furthermore, the delays are prohibitive if brain regions are farther apart¹³ or if they communicate through multiple routes with different numbers of intermediate stages.

A study by Siegel et al.⁵⁷ may appear to provide evidence for CTC. The authors required monkeys to memorize two items during a delay. During this task, they recorded neuronal activity in the prefrontal cortex and compared decoding accuracy for the items in memory across successive phases of the gamma oscillation. Interestingly, decoding accuracy was ~7% better during the optimal phase than during the worst phase of the oscillation. The authors also observed a phase difference (corresponding to ~5 ms) between the two memory items and suggested that different phases might permit the readout of specific object information. At the same time, decoding with more than 90% of the maximum accuracy was possible for both items during all phases, which is not in line with the phase-specific signals proposed by CTC.

Recent results suggests that gamma coherence between brain areas might be the result of communication rather than a prerequisite.⁵⁸ Furthermore, the gamma oscillations exhibit a dependence on the nature of the visual stimulus that is not yet fully understood. Some stimuli, such as grating patterns and red stimuli, elicit strong gamma in the EEG of humans and monkeys, whereas other stimuli do not.^{59–63} If gamma oscillations matter for perception, one would predict that stimuli eliciting strong gamma oscillations have a special status, but such findings have, to my knowledge, not been described.

Studies investigating neuronal synchrony in the visual cortex of monkeys that reported about perceptual grouping provided a direct test of BBS^{51,64–66} but did not support it. In an example study,⁵¹ monkeys had to mentally trace a target curve that was connected to a fixation point to identify a larger red circle at its other end as a target for eye movement (Figure 4A). There was also a second distracting curve that the monkeys had to ignore. The curve-tracing task requires the grouping of contour elements of the target curve and their segregation from the distractor. Recordings were made from V1 neurons with receptive fields at different positions on the curves. A first observation was that gamma synchronization did not occur. The V1 neurons did synchronize their responses, but the width of the peaks in the cross-correlation functions was ~100 ms (Figure 4A), indicative of a brain rhythm with a frequency lower than gamma. Synchronization was weak, with correlation coefficients around 0.01, and, surprisingly, did not occur at all in one of the monkeys (Figure 4B). Finally, the strength of synchronization was similar, regardless of whether the V1 receptive fields fell on the same or on different curves (Figures 4A and 4B). Synchrony could even be dissociated from binding by creating stimuli for which grouping decreased synchrony. These results indicate that binding does not require synchronicity in area V1.

Given that gamma synchrony is a local phenomenon, researchers proposed that binding could result from lower-frequency alpha oscillations (8–12 Hz), which synchronize neurons that are farther apart.⁶⁹ However, there are several reasons why

it is unlikely that alpha oscillations play a role in binding. The first was discussed above: the strength of synchrony in V1 at lower frequencies does not reflect the grouping of contours (Figures 4A and 4B). Second, there is substantial variability in the strength and frequency of alpha oscillations across individuals. This variability partially reflects genetic factors, because the EEG of monozygotic twins is more similar than that of dizygotic twins (Figure 4C).^{67,70} Importantly, some humans do not have an alpha peak in the power spectrum (e.g., the upper-left twin in Figure 4C; also see Haegens et al.⁷¹), a finding that is presumably related to the results in one of the monkeys in Figure 4B, which did not exhibit synchrony at all. Although human participants without a clear alpha rhythm were not described further, it seems unlikely that they had binding problems, because people with binding problems have severe symptoms.⁷² The absence of alpha oscillations in a fraction of the population is problematic for theories^{73,74} suggesting that this rhythm is causally involved in any cognitive function. Furthermore, alpha oscillations are suppressed when visual stimuli are presented and also when subjects direct their attention to task-relevant stimuli.^{42,74,75} Hence, alpha oscillations are suppressed when scene perception requires attentional grouping processes, further undermining their role in binding.

COINCIDENCE DETECTION

A cornerstone of BBS and CTC is that synchrony enhances the influence of neurons on their postsynaptic targets.⁷⁶ The efficiency with which synchronous inputs are integrated by a postsynaptic neuron depends on the distance between the membrane potential and the firing threshold of a neuron. If the membrane potential is far from the threshold, synchronous inputs help to drive a neuron. However, synchrony is less important if the membrane potential is close to the firing threshold or if the membrane potential fluctuates spontaneously.⁷⁷ Hence, whether synchrony matters for the postsynaptic drive is an empirical question and may differ between brain regions.^{47,78}

Histed and Maunsell⁶⁸ used optogenetics to test the influence of synchronicity on the activity of neurons in V1 of mice (Figure 4D). They compared short and strong optogenetic excitatory light pulses with weaker light pulses of longer duration. A strong light pulse of 3 ms induced a strong, synchronous burst of activity, whereas a weaker light pulse of 100 ms caused only a small, asynchronous increase in the firing rate that was maintained for the duration of the optogenetic stimulus. Interestingly, the total number of induced action potentials did not depend on the duration of the pulse but only on the total light energy (in mJ/mm²), proportional to the total number of photons reaching the neurons (inset in Figure 4D). The ability of the mouse to detect highly synchronous bursts of V1 activity, which were induced by the short pulse, was the same as its ability to detect the same number of spikes dispersed over 100 ms (Figure 4E). Although the optogenetic activation of neurons is different from how the cells are driven by visual stimuli, the result suggests that synchronicity does not increase the efficiency of postsynaptic integration, at least for intervals up to 100 ms. Apparently, downstream brain regions can efficiently integrate inputs across intervals up to 100 ms. In additional experiments, the authors showed that pulsing light at gamma frequencies was also

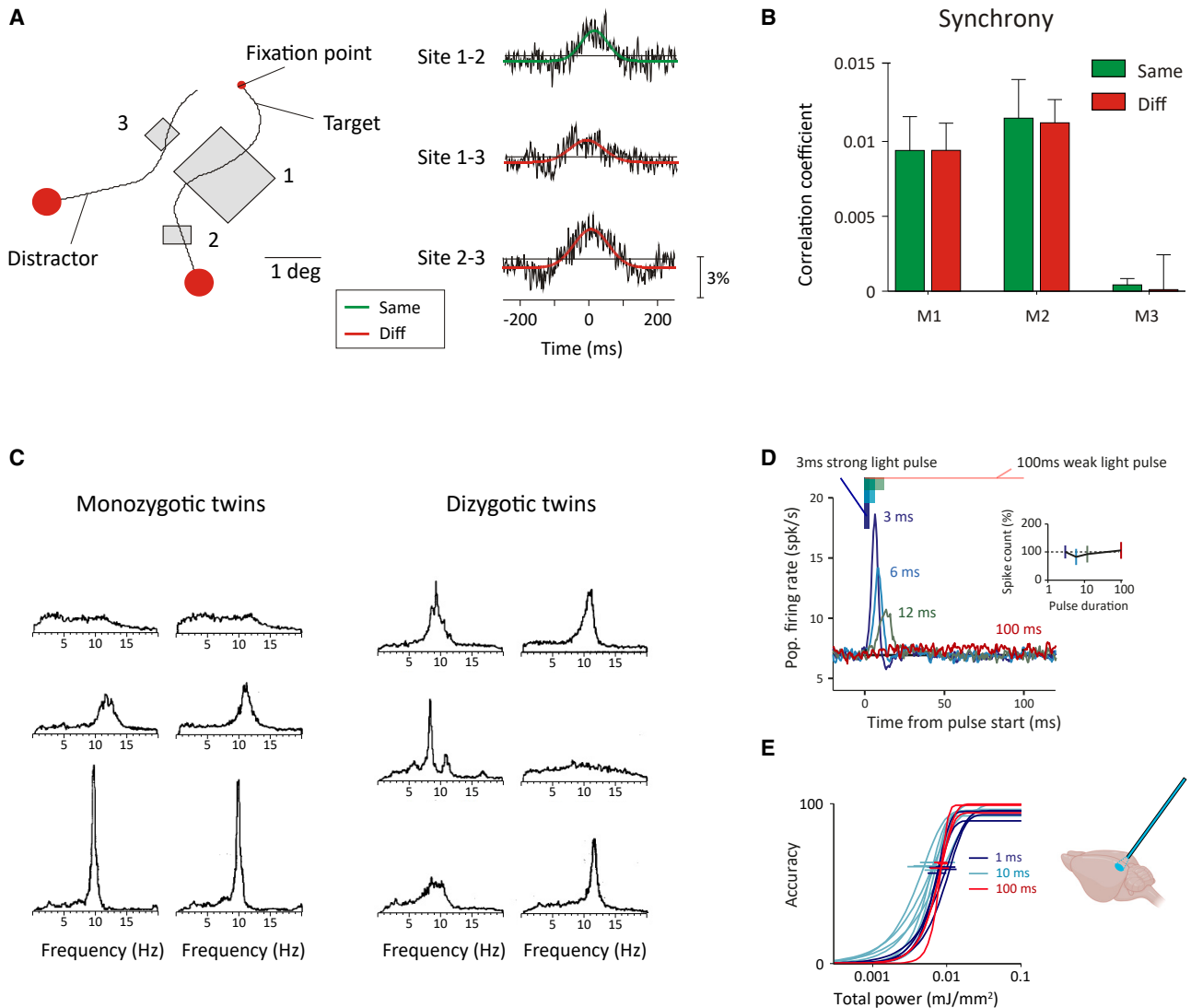


Figure 4. Synchrony, oscillations, and coincidence detection

(A) Monkeys traced a target curve that was connected to a fixation point to locate a larger red circle at the other end of this curve. They had to perceptually group the contours in the V1 receptive fields 1 and 2 and segregate them from contours of the distractor (receptive field 3) (left). Cross-correlation functions showing correlation coefficient as function of time lag. Peaks straddling the zero time bin indicate synchrony, which was similar for neurons coding for contours of the same curve (1–2) and neurons coding for different curves (1–3 and 2–3) (right).

(B) On average, the synchrony between neurons coding for contours of the same and different curves was similar. The largest differences were observed between monkeys, because V1 neurons in monkey M3 did not synchronize. Error bars denote S.E.M.

(C) EEG spectra of example monozygotic (left) and dizygotic twins (right).

(D) Optogenetic excitation of V1 neurons. V1 was stimulated with light pulses of different strengths and durations such that the light energy (in mJ) was kept constant. Inset, the number of elicited spikes was similar across light pulses with different durations. Error bars denote S.E.M.

(E) The ability of mice to detect V1 excitation was similar across light pulses of different durations, indicating that the degree of synchronicity did not matter for perception.

The data in (C) is from Lykken⁶⁷ and in (D) and (E) from Histed and Maunsell.⁶⁸

inconsequential for perception, suggesting that gamma has little impact on the information transfer between brain regions.

OBJECT-BASED ATTENTION: LABELING WITH AN ENHANCED FIRING RATE

Figure 5 illustrates the alternative hypothesis that features are bound in perception by labeling their representations across

brain regions with an enhanced firing rate (BBRE).^{15,16} BBRE proposes a dual role for firing rates. First, the firing rates reflect the tuning of neurons to features and base groupings during feedforward processing.^{16,79} Second, during incremental grouping, neurons coding for to-be-grouped features enhance their firing rate above the rate of neurons representing other objects and the background (Figure 5). At a psychological level of description, it is this “incremental representation” that receives

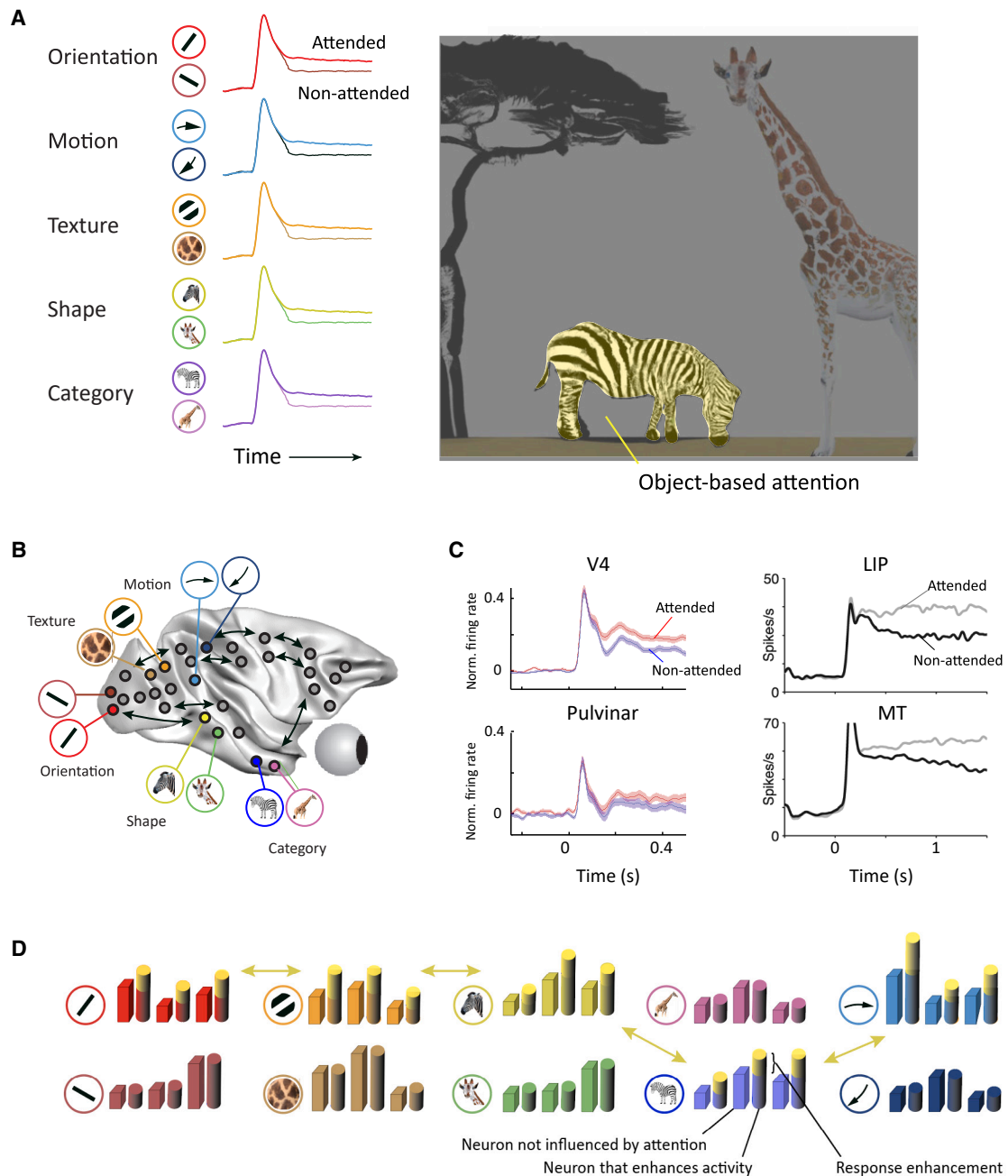


Figure 5. Labeling by firing rates as a mechanism for binding

(A) BBRE predicts that the activity of neurons coding for the features of an attended object is enhanced. Here, incremental groupings form for the zebra, which is segregated from the other objects and background (gray).

(B) Distributed representation of visual features.

(C) Attended objects elicit more activity than non-attended ones in V4, the pulvina of the thalamus, area LIP in parietal cortex, and motion-sensitive area MT. The data from V4 and the pulvina are from Zhou et al.,⁸¹ and the data from LIP and MT from Herrington and Assad.⁸²

(D) Features are represented across populations of neurons with various activity levels in each brain area (here, 6 neurons per area). Some neurons are influenced by attention (cylinders), whereas others are not (base representation, parallelepipeds). Yellow cylinders illustrate the response enhancement caused by incremental grouping. A pure attention signal can be extracted (and propagated) by comparing activity levels of neurons that are and are not influenced by incremental grouping (cylinders and parallelepipeds, next to each other). Yellow arrow, the attentional response enhancement needs to spread between brain regions.

object-based attention.^{16,80} The main prediction of BBRE is that when an object is attended, the representations of its features are enhanced across brain regions as an assembly and thereby grouped in perception.

In accordance with this prediction, attention increases the firing rates of neurons in most, if not all, cortical regions. Response enhancements occur in early visual cortex,^{83,84} mid-level visual areas such as V4 and MT (Figure 5C),^{81,85–87} inferotemporal cortex,^{88,89} parietal cortex,^{90–92} and frontal cortex^{93–97} (see Figure 5C for a few exemplary results). The typical finding is that the initial response elicited by feedforward processing is similar for attended and non-attended items and that the responses of attended objects are enhanced later, during a recurrent processing phase. Attention also increases firing rates in subcortical structures, including the pulvinar,^{81,98} the lateral geniculate nucleus (LGN),⁹⁹ and the basal ganglia.¹⁰⁰ Hence, when an object is attended, the increased neuronal responses elicited by its various features may integrate them into a coherent object representation. The co-selection of features of the same object in different brain regions implies that the distributed representations are linked by cortico-cortical and cortico-subcortical connections. Recent neural network models gained insight into the neuronal mechanisms for perceptual organization by equipping biologically realistic, deep neural networks with a feedback pathway.^{101,102} When an image is presented, the networks start with a feedforward sweep that detects object categories in the deeper levels of the network. They then select one of the categories in the upper layers and use the feedback pathway to enhance the relevant features at lower network levels (as in Figures 2B and 2C, for one category at a time). These studies illustrate how feedback connections between cortical areas could aid in image parsing by labeling relevant features with enhanced activity.

The co-selection of features of the same object plays a central role in various cognitive tasks, such as visual search and spatial cueing.¹⁰³ For example, during visual search, the target shape is kept in working memory as a template that enhances the activity of neurons in visual cortical areas that represent the location of the target object. The enhanced activity in retinotopic areas can be read out by other brain regions, linking the target shape to a location.^{104,105} Vice versa, in spatial cueing tasks, subjects look at an array of objects, and one of them is cued. The response enhancement at the cued location spreads through the recurrent connections to also enhance the representation of the shape so that it can be identified. Attentional operations like search and cueing can be chained flexibly into longer sequences to form more complex visual routines.^{106–109}

The idea that attention causes binding has a longer history. Treisman and her co-workers^{33,110} considered the role of feature binding in visual search. They noticed that visual search is slow when subjects must group features to detect the target of search, for example when they look for a red vertical line among green vertical and red horizontal lines. Treisman's feature integration theory (FIT) proposed that attention binds features into coherent object representations for one object at a time. In a search display, attention shifts serially from one object to the next to establish the feature conjunctions, which takes time. FIT's focus was on spatial attention, binding features at a location in space. Other studies demonstrated that it is also possible

to direct attention to one of two objects that overlap in space,^{15,17,18,111} implying that attentional selection can also be guided by feature dimensions other than space.

Furthermore, objects in images usually occupy larger regions so that a set of locations needs to be bound in perception (Figures 2A and 2B). For our interactions with objects, image segmentation is the process that groups the locations. In accordance with BBRE, image segmentation is associated with neuronal response enhancements in early visual brain regions, including the LGN, V1, V2, and V4.^{112–115} In the curve-tracing task, introduced above, V1 neurons representing contour elements of the target curve enhance their activity above the level elicited by a distractor curve (Figure 6A). The response enhancement thereby binds contour elements of the target curve into a coherent representation. At a psychological level of description, object-based attention is directed to the contour elements of the target curve,^{116,117} in accordance with BBRE.

V1 response enhancements also occur during figure-ground segregation, which occurs when the features of a figure differ from those of the background. In Figure 6B, for example, one perceives a square figure because the orientation of the image elements in this region differs from the orientation of the background elements.^{112,113,115,125} V1 activity elicited by figures is stronger than that elicited by the background, an effect known as figure-ground modulation (FGM). In area V1, the precision of FGM is high.^{113,119} The enhanced response is present throughout the figure, binding texture elements into a larger shape, but activity immediately drops for V1 neurons with receptive fields just outside the figure. FGM also occurs for more complex texture-defined shapes, as in Figure 6B, where the response enhancement labels all locations occupied by the "u" but is absent from the background between its legs.¹¹⁵ In accordance with BBRE, the enhanced neuronal activity spreads from attended image elements to other ones that group to them according to the Gestalt rules, such as color similarity, connectedness, and good continuation^{120,121} (Figures 6C and 6D), explaining the influence of low-level grouping cues on perceptual organization. Furthermore, BBRE occurs for natural images, where image elements of foreground objects elicit more activity than those that are part of the background (Figure 6E).^{122,123}

Kirchberger et al.¹²⁴ recently used optogenetics in mouse visual cortex to test the causal involvement of the recurrent labeling process in V1 in figure-ground perception. The mice either detected gratings on a gray background or saw texture-defined figures on a textured background. When optogenetics was used to block the entire response of V1 neurons, the mice had difficulties with both types of displays, as if they were largely blind (Figure 6F). However, when the optogenetic silencing was postponed to just after the feedforward response, the mice were able to perceive gratings on a gray background but unable to segregate figures from backgrounds. Hence, the late V1 response phase, which expresses FGM, is necessary for figure-ground perception and is read out by higher areas to plan a motor response.

Because firing rates code for both feature tuning and incremental grouping, there is a need to disentangle these influences. For example, differences in contrast could interfere with incremental grouping, because objects of high contrast elicit stronger firing rates than those with lower contrast.^{15,126} Importantly,

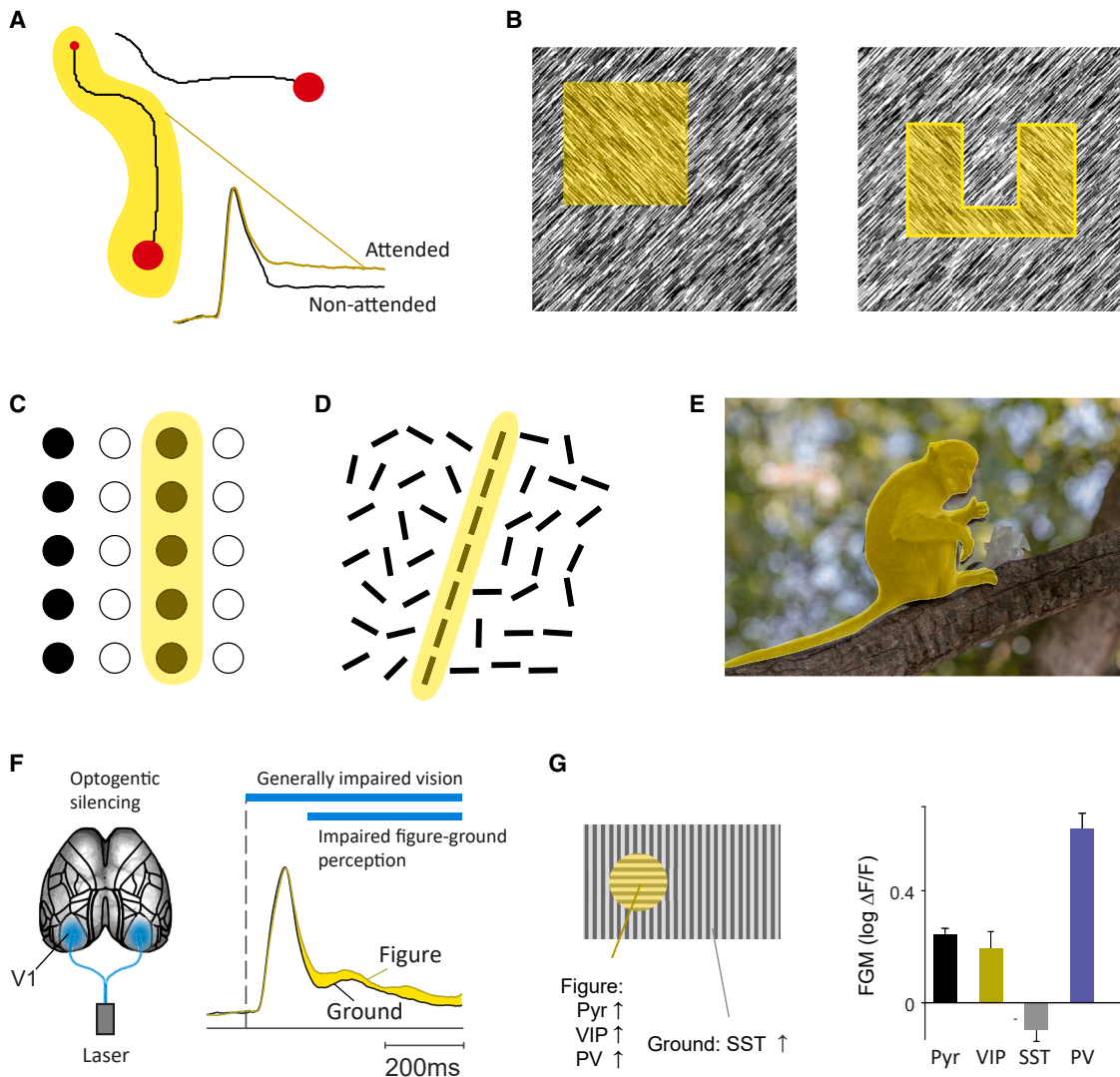


Figure 6. Incremental grouping of image elements at different spatial positions by enhanced firing rates

(A) In the curve-tracing task, the enhanced responses labels contour elements of the target curve, connected to the fixation point.^{83,118}
 (B) In texture-segregation tasks, the figure is defined by image elements that differ from the background. The V1 activity elicited by figural image elements is enhanced with a high spatial precision.^{112,113,115,119}
 (C) The attentional response enhancement spreads according to the Gestalt grouping rules. In this example, grouping occurs by color similarity.¹²⁰
 (D) A string of colinear image elements, standing out from other image elements with random orientation, is labeled in V1 with an enhanced response.¹²¹
 (E) Foreground objects in natural images elicit stronger neuronal responses than background image regions.^{122,123}
 (F) Optogenetic inhibition of V1 activity in both hemispheres of mice at different delays after stimulus appearance. The right side shows activity elicited by figure and ground in mouse V1. If optogenetic inhibition started shortly after stimulus onset, vision was generally impaired. If optogenetic inhibition was postponed to after the feedforward response, the mice could detect gratings displayed on a gray background, but they failed to perceive figure-ground stimuli.¹²⁴
 (G) Figures elicit stronger responses than the background in pyramidal neurons, VIP, and PV cells. FGM for SST neurons is reversed.¹²⁴ Error bars denote S.E.M.

some neurons are influenced by grouping and others are not (cylinders and parallelepipeds in Figure 5D). During the curve-tracing task (Figure 6A), for example, the V1 neurons that are little influenced by attention reliably represent the contrast of a curve. Other neurons are strongly modulated by attention so that pure contrast and attentional signals can be simultaneously decoded from the V1 population.¹²⁷ Similar results were obtained in mice performing the texture-segregation task. On average, figures elicit more activity of V1 pyramidal cells than backgrounds do, but there are also many pyramidal neurons without FGM.

Some of this variability is caused by differences between the layers of V1, because the response modulation is strongest in the superficial and deep layers and weakest in input layer 4.^{128,129}

Intriguingly, FGM also differs between pyramidal neurons and classes of interneurons (Figure 6G).¹²⁴ FGM is strongest for parvalbumin-positive (PV) neurons. This is followed by the FGM of pyramidal cells and vasoactive intestinal peptide-expressing (VIP) neurons. The VIP neurons disinhibit the cortical column,¹³⁰ and the FGM of pyramidal cells decreases if they

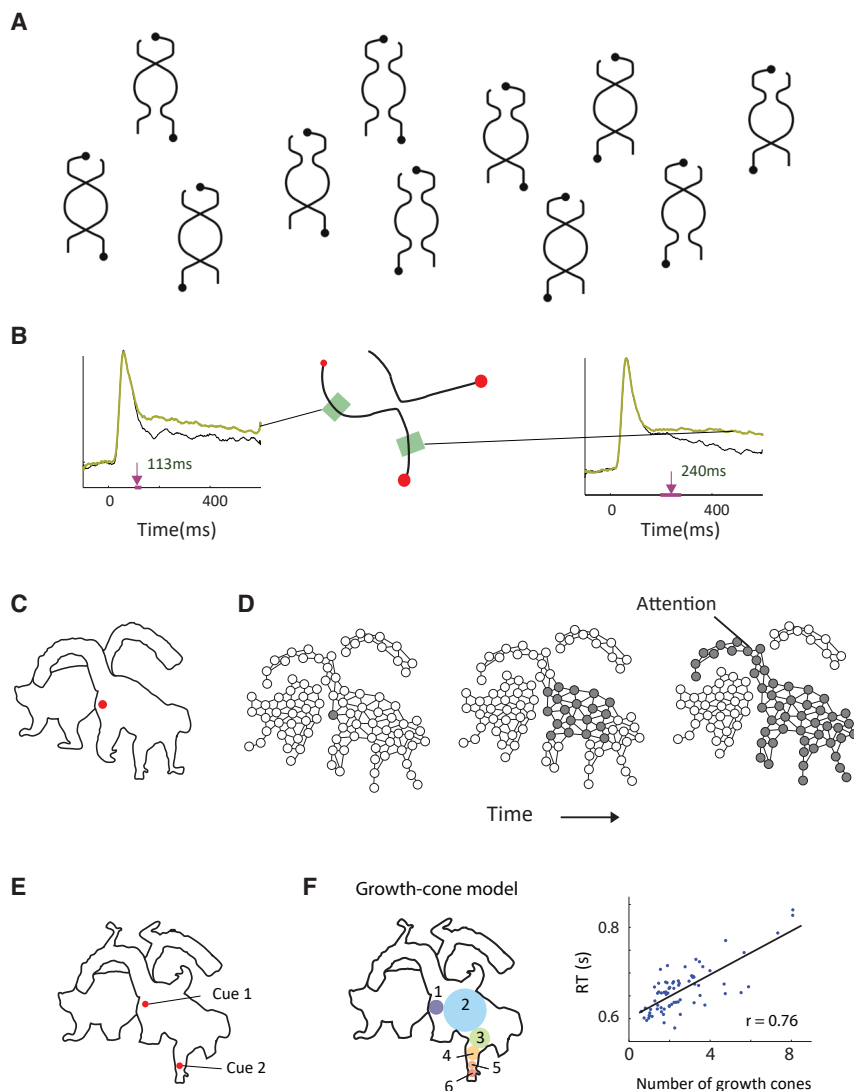


Figure 7. Time course and seriality of incremental grouping

(A) Try to find the curve that connects two circles. (B) The average V1 activity elicited by contour elements at the start of the target (yellow) or distractor curve (black) (left). Activity elicited by contour elements close to the end of the curve (right). (C) Incremental grouping of line drawings. (D) The incremental grouping process is hypothesized to fill the interior of one of the animals with enhanced activity. (E) Image parsing task. The subjects had to determine whether two cues fell on the same or on different objects. (F) The reaction time pattern was explained by a growth cone model, where the grouping process makes fast progress on large, homogeneous image regions. In the example of (E), cue 2 is reached after 6 growth cone shifts (left). The growth cone model predicts the reaction time across pictures with a correlation coefficient of 0.76 (right). (B) was modified from Pooremaeli and Roelfsema.¹¹⁸ (C)–(F) were adapted from Jeurissen et al.¹³³

PERCEPTUAL GROUPING IS TIME CONSUMING AND OCCURS FOR ONE OBJECT AT A TIME

An important difference between BBS and BBRE is the number of incremental groups that can coexist in perception (Figure 3B vs. Figure 5A). BBS permits several simultaneous incremental groups, labeled by unique patterns of synchrony. Because BBRE labels the incremental group with an enhanced activity level, a theoretical possibility is to define multiple firing rate levels to establish multiple simultaneous incremental groups. However, it is difficult to conceive of mechanisms that would enable the selective routing of the information from one of these groups, say, the one

are silenced.¹²⁴ Remarkably, FGM is reversed for somatostatin (SST) cells. Backgrounds elicit more activity of SST neurons than figures, in agreement with the hypothesis that these neurons help to suppress the representation of homogeneous image regions.¹³¹ It seems likely that the variations in the strength of the response enhancement across neurons, layers, and cell types can be exploited by neurons in other brain regions to separate the incremental grouping signal from other factors such as their feature tuning and stimulus contrast (Figure 5D). In accordance with this view, a recent study demonstrated that the interactions between V1 and V4 differ between the initial feedforward response and the later recurrent processing phase, as if there are separate “communication channels” between these neurons.¹³² The feedforward phase activates feature-selective neurons (both populations in Figure 5D), whereas the later, recurrent phase might correspond to the propagation of enhanced activity for incremental grouping (only cylinders in Figure 5D).

with second firing rate level from the top. Instead, theories of incremental grouping proposed that there is only one incremental group that is labeled with enhanced activity and thereby segregated from the rest.^{16,79,108} According to this view, features are either attended and grouped or they are not. An incremental group could, of course, be accompanied by several hardcoded feature conjunctions as base groupings.

To gain insight into the number of incremental groups, Houtkamp and Roelfsema⁷⁹ tested human participants in a curve-tracing task with multiple curves that they had to trace. Incremental grouping took place for only one curve at a time. Some intuition about the seriality of incremental grouping can be gleaned from Figure 7A, where the task is to find the curve connecting two circles (related to displays used by Houtkamp and Roelfsema⁷⁹).

In a typical curve-tracing task for monkeys illustrated in Figure 7B, the animal has to make an eye movement to the larger red circle that is connected by a target curve to a small

red fixation point. Recordings from area V1 revealed that the representation of the start of the curve is first enhanced before the increased activity spreads over contours farther along the curve.¹¹⁸ Hence, the curve's contour elements are grouped incrementally, a process that can be based on the Gestalt grouping cues of collinearity and connectedness. In human psychophysics, the processing time also increases with the length of the to-be-traced curve.¹³⁴ Interestingly, the tracing speed depends on the distance between the target curve and the distractors.¹³⁵ The speed is highest if there is a large distance between curves. These findings can be explained if the response enhancement spreads in multiple areas, like V1, V2, and V4.¹⁰⁸ In higher areas, neurons have larger receptive fields, and horizontal connections bridge large regions in the visual field so that tracing can make fast progress. However, if the curves are nearby, low-level areas, like V1, that represent the target curve at a higher resolution, with smaller receptive fields, may need to take over. In these lower areas, the horizontal connections link neurons with nearby receptive fields, which slows down the curve-tracing speed (see Poorasmaeili and Roelfsema¹¹⁸ for a discussion of the relation between the propagation speed of horizontal connections in the visual cortex and curve tracing). In accordance with the relation between the neuronal response enhancements and object-based attention, experiments in human participants revealed that object-based attention spreads gradually over the target curve.¹¹⁶

Incremental grouping also occurs when we parse natural images. The time to determine whether two cues are on the same object increases if they are farther apart or on different object parts.¹³⁶ Jeurissen et al.¹³³ measured processing delays while human subjects parsed line drawings (Figures 7C–7F). The subjects assessed whether two cues were on the same or different objects for line drawings of recognizable objects and modified line drawings in which the shapes could not be easily recognized (compare Figures 7C and 7E). The reaction time of the participants increased with the distance between the cues. Grouping was slower in narrow parts of the object and faster in wider parts, and the reaction time pattern was explained by a “growth cone” incremental grouping process in which the speed of grouping scales with the size of object parts. Specifically, the reaction time increased linearly with the minimal number of circular growth cones (colored circles in Figure 7F) that can be placed in the interior of the object to connect the two cues so that growth cone size depends on object width. The hypothesis is that an enhanced firing rate starts to spread at one of the cues within the object interior until the second cue is reached (Figure 7D), grouping labeled image regions. Larger growth cones could correspond to the receptive fields of neurons in higher areas and smaller growth cones to those in lower areas. Interestingly, the growth cone model explained less variance for recognizable objects than for the modified ones. This difference is presumably explained by semantic segmentation based on shape recognition in the inferotemporal cortex, which might only occur for recognizable objects. Shape-selective neurons could feed back to the lower areas and link features that are farther apart, like the head, tail, and legs of an animal (as in Figures 2B and 2C). Such shortcuts would explain the decrease

in the accuracy of models that only evaluate local, lower-level cues for incremental grouping.

Humans and monkeys make approximately three saccades per second. It is conceivable that the delays caused by image parsing, which can take up to hundreds of milliseconds, play a role in determining the typical duration of a fixation of 300 ms. Every visual fixation starts with a wave of feedforward processing lasting ~100 ms,¹⁹ which is followed by a recurrent processing phase for incremental grouping, lasting 100–300 ms. Studies that investigated the influence of saccades on image segmentation reported that they do not disrupt it, despite receptive field shifts in early visual areas. After every saccade, the response enhancements are routed back to the new set of neurons that now code for the contours of the attended object.^{137,138}

LINKING PERCEPTION TO ACTION

Let us now consider the role of incremental grouping in the visual guidance of actions. We direct our attention to the objects that we select for action, and some theoreticians argued that the distribution of visual attention is largely determined by action planning.^{139–141} Imagine trying to grasp one of the toys in the box of Figure 8A. Your fingers should be guided to the edges of one of the cars, and you should avoid the edges of other cars. Hence, grasping can rely on the response enhancements caused by incremental grouping, highlighting the shape of the relevant object in higher areas and its edges in low-level areas. Action selection in areas of the motor and frontal cortex can utilize the incremental groups, because they signal visual features that are informative for the transportation and shaping of the hand. It is a natural generalization of the incremental grouping process to also include (pre-)motor neurons coding for the action possibilities that an object provides, like whether it could be grasped, also known as affordances.²⁴ In line with the linkage of perception and action, attention is directed to items that become the target of an eye or hand movement. This coupling is so strong that observers are virtually unable to recognize shapes at locations other than the location that is the target of an upcoming eye movement.^{142,143}

Several studies measured the simultaneous enhancement of the representations of a visual object in the visual and frontal cortex.^{96,97,144} In one of these studies, Poorasmaeili et al.¹⁴⁴ recorded neuronal activity in areas V1, V4, and the frontal-eye fields (FEFs), an area involved in the generation of eye movements, during curve tracing (Figures 8B and 8C). The target curve elicited more activity than the distractor in all cortical areas, illustrating co-selection in visual and frontal cortex. Once the visual cortex propagated activity along the representation of the entire target curve, including the circle at its end, the frontal cortex can read out the enhanced response to plan the eye movement. Interestingly, the response enhancements in the visual and frontal cortex occurred virtually simultaneously. In a visual search task, Zhou and Desimone⁹⁶ also observed co-selection in the visual and frontal cortex but they found that the response enhancement in the FEFs preceded that in area V4 of visual cortex. A possible explanation for the delay during search is that the working memory of the target shape is maintained in the frontal cortex and fed back to the visual cortex

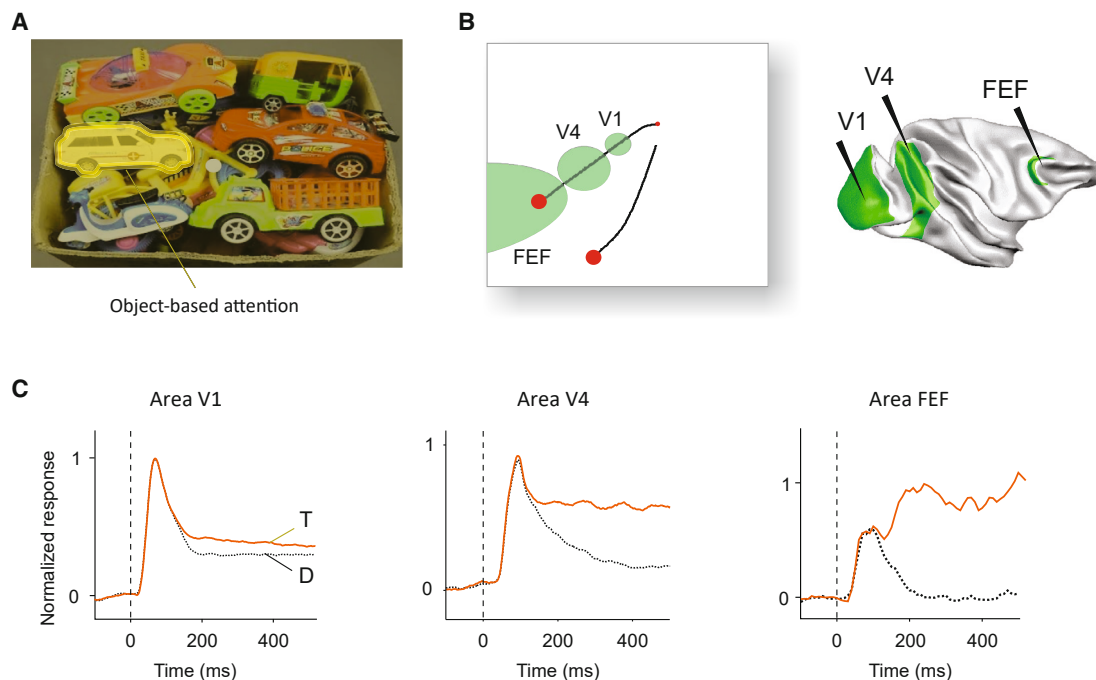


Figure 8. Interactions between the visual and motor system

(A) When grasping one of the toys in the box, it is essential to identify edges of the target toy and to segregate them from the edges of other objects.

(B) Curve-tracing stimulus with example receptive fields of neurons in area V1, V4, and the frontal-eye fields (FEFs).

(C) The response enhancement elicited by the target curve (T) over the response elicited by the distractor (D) was near simultaneous in areas of the visual and frontal cortex.

upon presentation of the search display. In contrast, the spread of the enhanced activity during curve tracing is guided by the shape of the curve itself, which is represented in visual cortex. Apparently, the relative timing of incremental grouping signals in different brain regions depends on the task and provides insight in the implementation of cognitive operations in the brain.

The precise connectivity patterns that ensure that features of the same objects are co-selected across brain regions remain to be fully understood. Moore and Armstrong¹⁴⁵ obtained important insights by directly testing the linkage of the attentional selection in the visual and frontal cortex. They activated neurons in the FEFs with electrical microstimulation and observed that this increased the activity of neurons in area V4 responding to the same object. Another study used microstimulation in combination with fMRI to demonstrate how FEF neurons influence activity throughout visual cortex.¹⁴⁶ Taken together, these experiments illustrate the strong links between representations in visual and frontal cortex.

The limitation of BBRE that only one incremental group can form at a time is expected to restrict the bandwidth of information transfer between the visual and frontal cortex. In contrast, BBS might permit multiple coexisting communication channels based on different synchronization patterns. Experiments in which subjects map multiple sensory stimuli onto different motor actions reveal dual-task interference, which is an inability to carry out two tasks at the same time.¹⁴⁷ When two stimuli are presented in rapid succession, the initial sensory processing of the two stimuli can proceed in parallel.¹⁴⁸ However, later pro-

cessing stages of the second stimulus must wait until these stages are freed up by the first stimulus. These results imply that there is a bottleneck in mapping stimuli into actions. Feed-forward processing of different sensory stimuli can occur in parallel, but the linkage of sensory and motor representations is limited to one incremental group, thereby causing the bottleneck,¹⁴⁹ which is in accordance with BBRE but not with BBS.

INCREMENTAL GROUPING AND THE GLOBAL NEURONAL WORKSPACE FOR CONSCIOUSNESS ACCESS

BBRE holds that response enhancements cause the co-selection of features of the same object across brain regions, providing building blocks for cognitive routines.^{103,150} This theoretical position is related to workspace theories of conscious perception.^{151–154} Workspace theories propose that perceptual objects reach awareness if their representations are amplified and broadcasted to processors across the brain. A recent study on the neuronal correlates of conscious access examined activity elicited by low-contrast visual stimuli in the visual and frontal cortex of monkeys.¹⁵⁵ The stimuli were close to the threshold of perception so that they were sometimes perceived and sometimes missed. The monkeys reported the stimuli only if they reached the frontal cortex at a sufficient strength and then elicited a sudden, strong, and sustained activity. This “ignition” process in the frontal cortex caused a top-down enhancement of activity in the visual cortex. I therefore propose that the global

neuronal workspace, in part, maps onto the neuronal mechanisms underlying object-based attention and that “broadcasting” may correspond to the propagation of enhanced activity for the co-selection of features of the same perceptual object across brain regions. This conjecture would explain why we consciously perceive integrated object representations rather than disconnected feature sets (recently reviewed by Mashour et al. ¹⁵¹).

Conclusions

In this review, I discussed evidence for BBS and BBRE, mechanisms that have been proposed to create coherent object representations from features represented by assemblies of neurons across different brain regions. Existing evidence does not support BBS. Gamma synchrony is too weak to bind neurons in different brain regions and too weak to link neurons in the same area that are separated by more than a few millimeters. Synchrony at lower frequencies is variable between subjects and is not related to binding. Furthermore, optogenetic experiments demonstrated that cortical neurons are equally driven by input that is coincident and dispersed across time intervals of up to 100 ms. Instead, the features of objects become bound in perception when their representations are labeled with an enhanced firing rate. This labeling process takes time and occurs for one object at a time. The assemblies of labeled neurons that form in the brain correspond to object-based attention in psychology. During action planning, the neuronal assemblies include neurons in frontal cortex representing affordances, explaining why efficient sensorimotor links form for one stimulus at a time.

Many questions about the binding problem remain open. For example, it is not yet clear which connectivity patterns within and between brain regions support the spread of enhanced activity to neurons that should become part of the same assembly. There also remains much to learn about the role of pyramidal cells and interneurons in the different layers of the cortex. Furthermore, many neuroscientists have focused on the cerebral cortex so that the contributions of interactions with subcortical structures such as the thalamus, basal ganglia, and the superior colliculus may not have been fully appreciated. Yet, I am optimistic that new insights are around the corner: the many tools of modern neuroscience, in combination with the deep-learning revolution in artificial intelligence, will help us to further shape our intuitions about what the neuronal networks in our brains are capable of.

ACKNOWLEDGMENTS

I thank Paolo Papale, Matthew Self, and Martin Vinck for their valuable comments on an earlier draft of the manuscript. This work was supported by NWO (Crossover Program 17619 “INTENSE” and 024.005.022 “DBI2,” a Gravitation program of the Dutch Ministry of Science, Education and Culture), Horizon Europe (ERC advanced grant 101052963 “NUMEROUS”; grant agreement 899287 “NeuraViper”), and the Human Brain Project (agreement no. 945539, “Human Brain Project SGA3”).

DECLARATION OF INTERESTS

Pieter Roelfsema is cofounder and shareholder of a neurotechnology start-up, Phosphoenix (Netherlands).

REFERENCES

1. Sincich, L.C., and Horton, J.C. (2005). The circuitry of V1 and V2: integration of color, form, and motion. *Annu. Rev. Neurosci.* *28*, 303–326.
2. Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. *Nature* *341*, 52–54.
3. Pasupathy, A., and Connor, C.E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* *86*, 2505–2519.
4. Nandy, A.S., Sharpee, T.O., Reynolds, J.H., and Mitchell, J.F. (2013). The Fine Structure of Shape Tuning in Area V4. *Neuron* *78*, 1102–1115. <https://doi.org/10.1016/j.neuron.2013.04.016>.
5. Conway, B.R., Moeller, S., and Tsao, D.Y. (2007). Specialized color modules in macaque extrastriate cortex. *Neuron* *56*, 560–573. <https://doi.org/10.1016/j.neuron.2007.10.008>.
6. Bashivan, P., Kar, K., and DiCarlo, J.J. (2019). Neural population control via deep image synthesis. *Science* *364*, eaav9436. <https://doi.org/10.1126/science.aav9436>.
7. Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* *1*, 1–47.
8. Kobatake, E., and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* *71*, 856–867.
9. Quiroga, R.Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nat. Rev. Neurosci.* *13*, 587–597.
10. Bausch, M., Niediek, J., Reber, T.P., Mackay, S., Boström, J., Elger, C.E., and Mormann, F. (2021). Concept neurons in the human medial temporal lobe flexibly represent abstract relations between concepts. *Nat. Commun.* *12*, 1–12. <https://doi.org/10.1038/s41467-021-26327-3>.
11. Singer, W., and Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* *18*, 555–586.
12. Singer, W., Engel, A.K., Kreiter, A.K., Munk, M.H., Neuenschwander, S., and Roelfsema, P.R. (1997). Neuronal assemblies: necessity, signature and detectability. *Trends Cogn. Sci.* *1*, 252–261.
13. Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron* *88*, 220–235. <https://doi.org/10.1016/j.neuron.2015.09.034>.
14. Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn. Sci.* *9*, 474–480. <https://doi.org/10.1016/j.tics.2005.08.011>.
15. Roelfsema, P.R., and Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Percept. Atten. Percept. Psychophys.* *73*, 2542–2572.
16. Roelfsema, P.R. (2006). Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* *29*, 203–227. <https://doi.org/10.1146/annurev.neuro.29.051605.112939>.
17. Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* *113*, 501–517.
18. Scholl, B.J. (2001). Objects and attention: the state of the art. *Cognition* *80*, 1–46.
19. Lamme, V.A., and Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* *23*, 571–579. [https://doi.org/10.1016/S0166-2236\(00\)01657-X](https://doi.org/10.1016/S0166-2236(00)01657-X).
20. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, eds. (Curran Associates Inc.), pp. 1–9.
21. Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* *19*, 356–365. <https://doi.org/10.1038/nn.4244>.

22. Richards, B.A., Lillicrap, T.P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R.P., de Berker, A., Ganguli, S., et al. (2019). A deep learning framework for neuroscience. *Nat. Neurosci.* **22**, 1761–1770. <https://doi.org/10.1038/s41593-019-0520-2>.
23. Nassi, J.J., and Callaway, E.M. (2011). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* **10**, 360–372.
24. Pezzulo, G., and Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends Cogn. Sci.* **20**, 414–424.
25. Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* **381**, 520–522.
26. Kar, K., Kubilius, J., Schmidt, K., Issa, E.B., and DiCarlo, J.J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* **22**, 974–983. <https://doi.org/10.1038/s41593-019-0392-5>.
27. Holcombe, A.O., and Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nat. Neurosci.* **4**, 127–128.
28. Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychol. Forsch.* **4**, 301–350.
29. Palmer, S.E. (1992). Common region: a new principle of perceptual grouping. *Cogn. Psychol.* **24**, 436–447.
30. Koffka, K. (1935). *Principles of Gestalt Psychology* (Routledge).
31. Li, H., Xiong, P., An, J., and Wang, L. (2019). Pyramid attention network for semantic segmentation. In *British Machine Vision Conference 2018*, pp. 1–13.
32. Hafiz, A.M., and Bhat, G.M. (2020). A survey on instance segmentation: state of the art. *Int. J. Multimed. Inf. Retr.* **9**, 171–189. <https://doi.org/10.1007/s13735-020-00195-x>.
33. Treisman, A., and Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cogn. Psychol.* **14**, 107–141.
34. von der Malsburg, C., and Schneider, W. (1986). A neural cocktail-party processor. *Biol. Cybern.* **54**, 29–40.
35. Engel, A.K., König, P., Kreiter, A.K., Schillen, T.B., and Singer, W. (1992). Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends Neurosci.* **15**, 218–226.
36. Gray, C.M., König, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337.
37. Castelo-Branco, M., Goebel, R., Neuenschwander, S., and Singer, W. (2000). Neural synchrony correlates with surface segregation rules. *Nature* **405**, 685–689.
38. Fries, P., Roelfsema, P.R., Engel, A.K., König, P., and Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc. Natl. Acad. Sci. USA* **94**, 12699–12704. <https://doi.org/10.1073/pnas.94.23.12699>.
39. Fries, P., Reynolds, J.H., Rorie, A.E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291**, 1560–1563.
40. Steinmetz, P.N., Roy, A., Fitzgerald, P.J., Hsiao, S.S., Johnson, K.O., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* **404**, 187–190.
41. Gregoriou, G.G., Gotts, S.J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* **324**, 1207–1210.
42. van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., and Roelfsema, P.R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. USA* **111**, 14332–14341. <https://doi.org/10.1073/pnas.1402773111>.
43. Ray, S., and Maunsell, J.H.R. (2015). Do gamma oscillations play a role in cerebral cortex? *Trends Cogn. Sci.* **19**, 78–85. <https://doi.org/10.1016/j.tics.2014.12.002>.
44. Lumer, E.D., Edelman, G.M., and Tononi, G. (1997). Neural dynamics in a model of the thalamocortical system. I. Layers, loops and the emergence of fast synchronous rhythms. *Cereb. Cortex* **7**, 207–227.
45. Sporns, O., Tononi, G., and Edelman, G.M. (1991). Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. *Proc. Natl. Acad. Sci. USA* **88**, 129–133.
46. Shastri, L., and Aijjanagadde, V. (1993). From simple associations to systematic reasoning: a connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behav. Brain Sci.* **16**, 417–451.
47. Shadlen, M.N., and Movshon, J.A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* **24**, 67–77.
48. Reynolds, J.H., and Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron* **24**, 19–29.
49. Engel, A.K., König, P., Gray, C.M., and Singer, W. (1990). Stimulus-dependent neuronal oscillations in cat visual cortex: Inter-columnar interaction as determined by cross-correlation analysis. *Eur. J. Neurosci.* **2**, 588–606.
50. Frien, A., and Eckhorn, R. (2000). Functional coupling shows stronger stimulus dependency for fast oscillations than for low-frequency components in striate cortex of awake monkey. *Eur. J. Neurosci.* **12**, 1466–1478.
51. Roelfsema, P.R., Lamme, V.A.F., and Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nat. Neurosci.* **7**, 982–991.
52. Nowak, L.G., Munk, M.H., James, A.C., Girard, P., and Bullier, J. (1999). Cross-correlation study of the temporal interactions between areas V1 and V2 of the macaque monkey. *J. Neurophysiol.* **81**, 1057–1074.
53. Bosman, C.A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., and Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron* **75**, 875–888.
54. Roberts, M.J., Lowet, E., Brunet, N.M., Ter Wal, M., Tiesinga, P., Fries, P., and De Weerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. *Neuron* **78**, 523–536.
55. Jia, X., Tanabe, S., and Kohn, A. (2013). Gamma and the Coordination of Spiking Activity in Early Visual Cortex. *Neuron* **77**, 762–774. <https://doi.org/10.1016/j.neuron.2012.12.036>.
56. Zandvakili, A., and Kohn, A. (2015). Coordinated neuronal activity enhances corticocortical communication. *Neuron* **87**, 827–839. <https://doi.org/10.1016/j.neuron.2015.07.026>.
57. Siegel, M., Warden, M.R., and Miller, E.K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proc. Natl. Acad. Sci. USA* **106**, 21341–21346.
58. Schneider, M., Broggin, A.C., Dann, B., Tzanou, A., Uran, C., Sheshadri, S., Scherberger, H., and Vinck, M. (2021). A mechanism for inter-areal coherence through communication based on connectivity and oscillatory power. *Neuron* **109**, 4050–4067.e12. <https://doi.org/10.1016/j.neuron.2021.09.037>.
59. Hermes, D., Miller, K.J., Wandell, B.A., and Winawer, J. (2015). Stimulus dependence of gamma oscillations in human visual cortex. *Cereb. Cortex* **25**, 2951–2959. <https://doi.org/10.1093/cercor/bhu091>.
60. Gieselmann, M.A., and Thiele, A. (2008). Comparison of spatial integration and surround suppression characteristics in spiking activity and the local field potential in macaque V1. *Eur. J. Neurosci.* **28**, 447–459.
61. Bartoli, E., Bosking, W., and Foster, B.L. (2020). Seeing Visual Gamma Oscillations in a New Light. *Trends Cogn. Sci.* **24**, 501–503. <https://doi.org/10.1016/j.tics.2020.03.009>.
62. Peter, A., Uran, C., Klon-Lipok, J., Roese, R., Van Stijn, S., Barnes, W., Dowdall, J.R., Singer, W., Fries, P., and Vinck, M. (2019). Surface color

- and predictability determine contextual modulation of V1 firing and gamma oscillations. *Elife* 8, e42101–e42138. <https://doi.org/10.7554/eLife.42101>.
63. Shirhatti, V., and Ray, S. (2018). Long-wavelength (reddish) hues induce unusually large gamma oscillations in the primate primary visual cortex. *Proc. Natl. Acad. Sci. USA* 115, 4489–4494. <https://doi.org/10.1073/pnas.1717334115>.
 64. Lamme, V.A.F., and Spekreijse, H. (1998). Neuronal synchrony does not represent texture segregation. *Nature* 396, 362–366.
 65. Thiele, A., and Stoner, G. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature* 421, 366–370.
 66. Palanca, B.J.A., and DeAngelis, G.C. (2005). Does neuronal synchrony underlie visual feature grouping? *Neuron* 46, 333–346.
 67. Lykken, D.T. (1982). Presidential address, 1981. Research with twins: the concept of emergence. *Psychophysiology* 19, 361–373. <https://doi.org/10.1111/j.1469-8986.1982.tb02489.x>.
 68. Histed, M.H., and Maunsell, J.H.R. (2014). Cortical neural populations can guide behavior by integrating inputs linearly, independent of synchrony. *Proc. Natl. Acad. Sci. USA* 111, E178–E187. <https://doi.org/10.1073/pnas.1318750111>.
 69. Zhang, Y., Zhang, Y., Cai, P., Luo, H., and Fang, F. (2019). The causal role of α -oscillations in feature binding. *Proc. Natl. Acad. Sci. USA* 116, 17023–17028. <https://doi.org/10.1073/pnas.1904160116>.
 70. Linkenkaer-Hansen, K., Smit, D.J.A., Barkil, A., Van Beijsterveldt, T.E.M., Brussaard, A.B., Boomsma, D.I., Van Ooyen, A., and De Geus, E.J.C. (2007). Genetic contributions to long-range temporal correlations in ongoing oscillations. *J. Neurosci.* 27, 13882–13889. <https://doi.org/10.1523/JNEUROSCI.3083-07.2007>.
 71. Haegens, S., Cousijn, H., Wallis, G., Harrison, P.J., and Nobre, A.C. (2014). Inter- and intra-individual variability in alpha peak frequency. *Neuroimage* 92, 46–55. <https://doi.org/10.1016/j.neuroimage.2014.01.049>.
 72. Robertson, L.C. (2003). Binding, spatial attention and perceptual awareness. *Nat. Rev. Neurosci.* 4, 93–102.
 73. Miller, E.K., Lundqvist, M., and Bastos, A.M. (2018). Working Memory 2.0. *Neuron* 100, 463–475. <https://doi.org/10.1016/j.neuron.2018.09.023>.
 74. Jensen, O., Gips, B., Bergmann, T.O., and Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends Neurosci.* 37, 357–369. <https://doi.org/10.1016/j.tins.2014.04.001>.
 75. Thut, G., Nietzel, A., Brandt, S.A., and Pascual-Leone, A. (2006). α -Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.* 26, 9494–9502. <https://doi.org/10.1523/JNEUROSCI.0875-06.2006>.
 76. König, P., Engel, A.K., and Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci.* 19, 130–137.
 77. Shadlen, M.N., and Newsome, W.T. (1994). Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* 4, 569–579.
 78. Kempter, R., Gerstner, W., and Van Hemmen, J.L. (1998). How the threshold of a neuron determines its capacity for coincidence detection. *Biosystems* 48, 105–112. [https://doi.org/10.1016/S0303-2647\(98\)00055-0](https://doi.org/10.1016/S0303-2647(98)00055-0).
 79. Houtkamp, R., and Roelfsema, P.R. (2010). Parallel and serial grouping of image elements in visual perception. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1443–1459.
 80. Duncan, J., Humphreys, G., and Ward, R. (1997). Competitive brain activity in visual attention. *Curr. Opin. Neurobiol.* 7, 255–261.
 81. Zhou, H., Schafer, R.J., and Desimone, R. (2016). Pulvinar-cortex interactions in vision and attention. *Neuron* 89, 209–220. <https://doi.org/10.1016/j.neuron.2015.11.034>.
 82. Herrington, T.M., and Assad, J.A. (2010). Temporal sequence of attention modulation in the lateral intraparietal area and middle temporal area during rapid covert shifts of attention. *J. Neurosci.* 30, 3287–3296.
 83. Roelfsema, P.R., Lamme, V.A., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395, 376–381.
 84. Herrero, J.L., Roberts, M.J., Delicato, L.S., Gieselmann, M.A., Dayan, P., and Thiele, A. (2008). Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature* 454, 1110–1114.
 85. Herrington, T.M., and Assad, J.A. (2009). Neural activity in the middle temporal area and lateral intraparietal area during endogenously cued shifts of attention. *J. Neurosci.* 29, 14160–14176.
 86. Treue, S., and Maunsell, J.H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382, 539–541.
 87. Chelazzi, L., Miller, E.K., Duncan, J., and Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* 11, 761–772.
 88. Chelazzi, L., Miller, E.K., Duncan, J., and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature* 363, 345–347.
 89. Stemmann, H., and Freiwald, W.A. (2019). Evidence for an attentional priority map in inferotemporal cortex. *Proc. Natl. Acad. Sci. USA* 116, 23797–23805. <https://doi.org/10.1073/pnas.1821866116>.
 90. Bisley, J.W., and Goldberg, M.E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science* 299, 81–86.
 91. Constantinidis, C., and Steinmetz, M.A. (2001). Neuronal responses in area 7a to multiple-stimulus displays: I. Neurons encode the location of the salient stimulus. *Cereb. Cortex* 11, 581–591.
 92. Yokoi, I., and Komatsu, H. (2009). Relation between neural responses and visual grouping in the monkey parietal cortex. *J. Neurosci.* 29, 13210–13221.
 93. Schall, J.D., and Thompson, K.G. (1999). Neural selection and control of visually guided eye movements. *Annu. Rev. Neurosci.* 22, 241–259.
 94. Everling, S., Tinsley, C.J., Gaffan, D., and Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nat. Neurosci.* 5, 671–676.
 95. Buschman, T.J., and Miller, E.K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862.
 96. Zhou, H., and Desimone, R. (2011). Feature-based attention in the frontal eye field and area V4 during visual search. *Neuron* 70, 1205–1217.
 97. Bichot, N.P., Heard, M.T., Degennaro, E.M., and Desimone, R. (2015). A source for feature-based attention in the prefrontal cortex. *Neuron* 88, 832–844. <https://doi.org/10.1016/j.neuron.2015.10.001>.
 98. Saalmann, Y.B., Pinsk, M.A., Wang, L., Li, X., and Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337, 753–756.
 99. McAlonan, K., Cavanaugh, J., and Wurtz, R.H. (2008). Guarding the gateway to cortex with attention in the visual thalamus. *Nature* 456, 391–394.
 100. Boussaoud, D., and Kermadi, I. (1997). The primate striatum: neuronal activity in relation to spatial attention versus motor preparation. *Eur. J. Neurosci.* 9, 2152–2168.
 101. Lei, J., Benjamin, A.S., and Kording, K.P. (2021). Object based attention through internal gating. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2106.04540>.
 102. Papale, P., Williford, J.R., Balk, S., and Roelfsema, P.R. (2023). Modulatory feedback explain object segmentation by attention. Preprint at bioRxiv. <https://doi.org/10.1101/2023.01.19.524712>.
 103. Roelfsema, P.R., Lamme, V.A., and Spekreijse, H. (2000). The implementation of visual routines. *Vision Res.* 40, 1385–1411.

104. Olivers, C.N.L., Peters, J., Houtkamp, R., and Roelfsema, P.R. (2011). Different states in visual working memory : when it guides attention and when it does not. *Trends Cogn. Sci.* **15**, 327–334. <https://doi.org/10.1016/j.tics.2011.05.004>.
105. Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222.
106. Moro, S.I., Tolboom, M., Khayat, P.S., and Roelfsema, P.R. (2010). Neuronal activity in the visual cortex reveals the temporal order of cognitive operations. *J. Neurosci.* **30**, 16293–16303. <https://doi.org/10.1523/JNEUROSCI.1256-10.2010>.
107. Roelfsema, P.R., Khayat, P.S., and Spekreijse, H. (2003). Subtask sequencing in the primary visual cortex. *Proc. Natl. Acad. Sci. USA* **100**, 5467–5472. <https://doi.org/10.1073/pnas.0431051100>.
108. Roelfsema, P.R., and de Lange, F.P. (2016). Early visual cortex as a multi-scale cognitive blackboard. *Annu. Rev. Vis. Sci.* **2**, 131–151. <https://doi.org/10.1146/annurev-vision-111815-114443>.
109. Ullman, S. (1984). Visual routines. *Cognition* **18**, 97–159.
110. Treisman, A.M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136.
111. O’Craven, K.M., Downing, P.E., and Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature* **401**, 584–587.
112. Lamme, V.A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* **15**, 1605–1615.
113. Poort, J., Raudies, F., Wannig, A., Lamme, V.A.F., Neumann, H., and Roelfsema, P.R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* **75**, 143–156.
114. Lorteije, J.A.M., Zylberberg, A., Ouellette, B.G., de Zeeuw, C.I., Sigman, M., and Roelfsema, P.R. (2015). The formation of hierarchical decisions in the visual cortex. *Neuron* **87**, 1344–1356. <https://doi.org/10.1016/j.neuron.2015.08.015>.
115. Self, M.W., Jeurissen, D., van Ham, A.F., van Vugt, B., Poort, J., and Roelfsema, P.R. (2019). The segmentation of proto-objects in the monkey primary visual cortex. *Curr. Biol.* **29**, 1019–1029.e4. <https://doi.org/10.1016/j.cub.2019.02.016>.
116. Houtkamp, R., Spekreijse, H., and Roelfsema, P.R. (2003). A gradual spread of attention during mental curve tracing. *Percept. Psychophys.* **65**, 1136–1144.
117. Scholte, H.S., Spekreijse, H., and Roelfsema, P.R. (2001). The spatial profile of visual attention in mental curve tracing. *Vision Res.* **41**, 2569–2580.
118. Pooresmaeili, A., and Roelfsema, P.R. (2014). A growth-cone model for the spread of object-based attention during contour grouping. *Curr. Biol.* **24**, 2869–2877. <https://doi.org/10.1016/j.cub.2014.10.007>.
119. Lamme, V.A., Rodriguez-Rodriguez, V., and Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cereb. Cortex* **9**, 406–413.
120. Wannig, A., Stanisor, L., and Roelfsema, P.R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nat. Neurosci.* **14**, 1243–1244.
121. Li, W., Piëch, V., and Gilbert, C.D. (2006). Contour saliency in primary visual cortex. *Neuron* **50**, 951–962.
122. Papale, P., Leo, A., Cecchetti, L., Handjaras, G., Kay, K.N., Pietrini, P., and Ricciardi, E. (2018). Foreground-background segmentation revealed during natural image viewing. *eNeuro* **5**, *In press*. <https://doi.org/10.1523/ENEURO.0075-18.2018>.
123. Papale, P., Zuiderbaan, W., Teeuwen, R.R.M., Gilhuis, A., Self, M.W., Roelfsema, P.R., and Dumoulin, S.O. (2021). The influence of objecthood on the representation of natural images in the visual cortex. Preprint at bioRxiv. e0075-18. <https://doi.org/10.1101/2021.09.21.461209>.
124. Kirchberger, L., Mukherjee, S., Schnabel, U.H., van Beest, E.H., Barszegyan, A., Levelt, C.N., Heimel, J.A., Lorteije, J.A.M., van der Togt, C., Self, M.W., and Roelfsema, P.R. (2021). The essential role of recurrent processing for figure-ground perception in mice. *Sci. Adv.* **7**, eabe1833. <https://doi.org/10.1126/sciadv.abe1833>.
125. Zipser, K., Lamme, V.A., and Schiller, P.H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* **16**, 7376–7389.
126. Watson, P., Korjoukov, I., Vartak, D., and Roelfsema, P.R. (2013). Luminance contrast has little influence on the spread of object-based attention. *Vision Res.* **85**, 90–103. <https://doi.org/10.1016/j.visres.2012.12.010>.
127. Pooresmaeili, A., Poort, J., Thiele, A., and Roelfsema, P.R. (2010). Separable codes for attention and luminance contrast in the primary visual cortex. *J. Neurosci.* **30**, 12701–12711.
128. Self, M.W., van Kerkoerle, T., Supèr, H., and Roelfsema, P.R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Curr. Biol.* **23**, 2121–2129.
129. van Kerkoerle, T., Self, M.W., and Roelfsema, P.R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat. Commun.* **8**, 13804–13812. <https://doi.org/10.1038/ncomms13804>.
130. Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J.I., Huang, Z.J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature* **503**, 521–524.
131. Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z.J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature* **490**, 226–231.
132. Semedo, J.D., Jasper, A.I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C.K., Kohn, A., and Yu, B.M. (2022). Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nat. Commun.* **13**, 1–14. <https://doi.org/10.1038/s41467-022-28552-w>.
133. Jeurissen, D., Self, M.W., and Roelfsema, P.R. (2016). Serial grouping of 2D-image regions with object-based attention in humans. *Elife* **5**, e14320. <https://doi.org/10.7554/eLife.14320>.
134. Jolicoeur, P., Ullman, S., and MacKay, M. (1986). Curve tracing: a possible basic operation in the perception of spatial relations. *Mem. Cognit.* **14**, 129–140.
135. Jolicoeur, P., and Ingleton, M. (1991). Size invariance in curve tracing. *Mem. Cognit.* **19**, 21–36.
136. Korjoukov, I., Jeurissen, D., Kloosterman, N.A., Verhoeven, J.E., Scholte, H.S., and Roelfsema, P.R. (2012). The time course of perceptual grouping in natural scenes. *Psychol. Sci.* **23**, 1482–1489. <https://doi.org/10.1177/0956797612443832>.
137. Khayat, P.S., Spekreijse, H., and Roelfsema, P.R. (2004). Correlates of transaccadic integration in the primary visual cortex of the monkey. *Proc. Natl. Acad. Sci. USA* **101**, 12712–12717. <https://doi.org/10.1073/pnas.0301935101>.
138. O’Herron, P., and von der Heydt, R. (2009). Short-term memory for figure-ground organization in the visual cortex. *Neuron* **61**, 801–809.
139. Allport, A. (1989). *Visual Attention*, pp. 631–682.
140. Baldauf, D., and Deubel, H. (2010). Attentional landscapes in reaching and grasping. *Vision Res.* **50**, 999–1013. <https://doi.org/10.1016/j.visres.2010.02.008>.
141. Sheliga, B.M., Riggio, L., and Rizzolatti, G. (1994). Orienting of attention and eye movements. *Exp. Brain Res.* **98**, 507–522.
142. Kowler, E., Anderson, E., Doshier, B., and Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Res.* **35**, 1897–1916.
143. Deubel, H., and Schneider, W.X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Res.* **36**, 1827–1837.
144. Pooresmaeili, A., Poort, J., and Roelfsema, P.R. (2014). Simultaneous selection by object-based attention in visual and frontal cortex. *Proc.*

- Natl. Acad. Sci. USA *111*, 6467–6472. <https://doi.org/10.1073/pnas.1316181111>.
145. Moore, T., and Armstrong, K.M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature* *421*, 370–373.
146. Ekstrom, L.B., Roelfsema, P.R., Arsenault, J.T., Bonmassar, G., and Vanduffel, W. (2008). Bottom-up dependent gating of frontal signals in early visual cortex. *Science* *321*, 414–417.
147. Pashler, H. (1994). Dual-Task Interference in Simple Tasks: Data and Theory. *Psychol. Bull.* *116*, 220–244. <https://doi.org/10.1037//0033-2909.116.2.220>.
148. Sigman, M., and Dehaene, S. (2005). Parsing a cognitive task: a characterization of the mind's bottleneck. *PLoS Biol.* *3*, e37.
149. Zylberberg, A., Fernández Slezak, D., Roelfsema, P.R., Dehaene, S., and Sigman, M. (2010). The brain's router: a cortical network model of serial processing in the primate brain. *PLoS Comput. Biol.* *6*, e1000765.
150. Roelfsema, P.R. (2005). Elemental operations in vision. *Trends Cogn. Sci.* *9*, 226–233. <https://doi.org/10.1016/j.tics.2005.03.012>.
151. Mashour, G.A., Roelfsema, P., Changeux, J.P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron* *105*, 776–798. <https://doi.org/10.1016/j.neuron.2020.01.026>.
152. Dehaene, S., and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* *70*, 200–227. <https://doi.org/10.1016/j.neuron.2011.03.018>.
153. Dehaene, S., Kerszberg, M., and Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA* *95*, 14529–14534.
154. Baars, B.J., and Franklin, S. (2003). How conscious experience and working memory interact. *Trends Cogn. Sci.* *7*, 166–172. [https://doi.org/10.1016/S1364-6613\(03\)00056-1](https://doi.org/10.1016/S1364-6613(03)00056-1).
155. van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P.R. (2018). The threshold for conscious report: signal loss and response bias in visual and frontal cortex. *Science* *360*, 537–542. <https://doi.org/10.1126/science.aar7186>.