UNIVERSITY OF RIJEKA

POSTGRADUATE DOCTORAL STUDIES IN PSYCHOLOGY

Mateja Marić

NEURODYNAMIC MODELS OF TOP-DOWN EFFECTS ON VISUAL PERCEPTION

DOCTORAL THESIS

Rijeka, 2021

UNIVERSITY OF RIJEKA
POSTGRADUATE DOCTORAL STUDIES IN PSYCHOLOGY

Mateja Marić

NEURODYNAMIC MODELS OF TOP-DOWN EFFECTS ON VISUAL PERCEPTION

DOCTORAL THESIS

Rijeka, 2021

UNIVERSITY OF RIJEKA
POSTGRADUATE DOCTORAL STUDIES IN PSYCHOLOGY

Mateja Marić

NEURODYNAMIC MODELS OF TOP-DOWN EFFECTS ON VISUAL PERCEPTION

DOCTORAL THESIS

Mentor: prof. dr. sc. Dražen Domijan

Rijeka, 2021

This page is intentionally left blank

SVEUČILIŠTE U RIJECI

POSLIJEDIPLOMSKI SVEUČILIŠNI DOKTORSKI STUDIJ IZ PSIHOLOGIJE

Mateja Marić

NEURODINAMIČKI MODELI SILAZNIH PROCESA U VIDNOJ PERCEPCIJI

DOKTORSKI RAD

Mentor: prof. dr. sc. Dražen Domijan

Rijeka, 2021

Thesis mentor: prof. dr. sc. Dražen Domijan,

Faculty of Humanities and Social Sciences, University of Rijeka

The doctoral thesis was defended on ___________ at the Faculty of Humanities and Social Sciences at the University of Rijeka, in front of the examining committee consisted of:

1. izv. prof. dr. sc. Igor Bajšanski (committee chairman)
2. prof. dr. sc. Mladenka Tkalčić (committee member)
3. prof. dr. sc. Pavle Valerjev (committee member)

ACKNOWLEDGEMENT

The doctoral thesis was supported by the Croatian Science Foundation Research Grant HRZZ-IP-11-2013-4139 and the University of Rijeka Grant 13.04.1.3.11.

I would like to thank my mentor prof. dr. sc. Dražen Domijan, without whom I would have never been introduced with the field of theoretical and computational neuroscience. Thank you for helping me find my professional path and interests! I am very grateful for your patience, motivation, accessibility, and for "all those skills that one mentor should have".

Many thanks to my parents Mato and Ankica, who provided me with the first support for enrolling in the doctoral studies. Thank you for all your encouraging messages, advice, and understanding.

The most of all, thanks to my husband Igor and son Jakov who were always beside me and sacrificed our family time to support me to accomplish my professional goals. Thank you for going above and beyond!

ABSTRACT

The aim of the doctoral thesis is to develop new neural network models that will explore how feedback projections in the visual cortex contribute to the top-down modulations of visual perception. Two types of top-down effects are considered: 1) visual selective attention and 2) prior expectations. The models represent modifications and extensions of previously published models of lateral inhibition and the adaptive resonance theory. The proposed models are thoroughly evaluated using computer simulations implemented in the MATLAB. Model’s output is compared with behavioral and neural data.

In the first part of the thesis, a model of recurrent competitive network has been developed with the ability to flexibly orient attention in a spatial map to either: a single location in space, to all locations occupied by an object, or to all locations occupied by the feature value. To achieve this property, the network was augmented by biophysically plausible mechanisms emulating properties of synaptic and dendritic computation. If the proposed network is further embedded in a larger multi-scale neural architecture for boundary detection, it was able to simulate object-based attention and implement visual routines such as mental contour tracing.

In the second part of the thesis, a neural network for color perception based on the adaptive resonance theory has been developed. The model explains how feedback projections contribute to stable learning of color codes and conscious experience of colors. The model shows that the same mechanisms that assure stability of learning are also responsible for constraining the effect of top-down expectations on color perception. In general, the model shows that top-down predictions, to a large extent, do not alter the content of conscious visual perception.

Keywords: neural networks, visual attention, top-down effects, cognitive impenetrability of vision, adaptive resonance theory, color perception

PROŠIRENI SAŽETAK

Neuroznanstvena istraživanja pokazuju da se vidna percepcija odvija u nizu hijerarhijski organiziranih kortikalnih mreža među kojima postoje uzlazne, lateralne i povratne veze. Postojanje povratnih veza upućuje na zaključak da silazni procesi kao što su očekivanja i uvjerenja nastala na osnovi prethodno stečenog znanja mogu mijenjati sadržaj perceptivnog iskustva i time direktno utjecati na vidnu percepciju. Opći cilj i svrha doktorskog rada je razvoj novih matematičkih modela neuronskih mreža i njihovo testiranje putem računalnih simulacija. Modeli su dizajnirani s ciljem pružanja uvida u to kako povratne veze u vidnom korteksu doprinose silaznim utjecajima na vidnu percepciju. U radu su razmotrene dvije vrste silaznih utjecaja: 1) vidna selektivna pažnja i 2) očekivanja. Novi modeli konstruirani su putem modifikacija postojećih modela lateralne inhibicije i teorije adaptivne rezonance kako bi objasnili recentne psihofizičke i neuroznanstvene nalaze o odnosu između selekcije lokacija, objekata i obilježja u prostoru i utjecaja očekivanja na vidnu percepciju. Metoda doktorskog istraživanja je simulacijsko modeliranje neuronskih mreža gdje se matematički opisi modela implementiraju i testiraju na računalu putem programa MATLAB. Ovom se metodom sustavno ispituju odgovori modela na podražaje slične onima koji se zadaju ispitanicima u psihološkim istraživanjima, a potom se dobiveni rezultati modela uspoređuju s empirijskim podacima.

U prvom dijelu doktorskog rada razvijena je rekurentna kompetitivna mreža s lateralnom inhibicijom za modeliranje selektivne pažnje. Na prostornoj mapi mreža ima sposobnost fleksibilnog odabira: jedne lokacije u prostoru, svih lokacija koje pripadaju nekom objektu ili svih lokacija koje pripadaju nekom obilježju. Mreža je proširena računalnim mehanizmima koji oponašaju moguću ulogu sinaptičkog prijenosa i dendrita u kortikalnoj obradi informacija. Drugim riječima, u mrežu su uključeni dendriti kao samostalne računalne jedinice te samoregulacija sinaptičkog prijenosa putem retrogradnih signala. Ovi mehanizmi povećavaju stabilnost mreže, smanjuju ukupnu količinu inhibicije u mreži i omogućavaju istovremenu selekciju arbitrarnog broja pobjednika koji su definirani zajedničkim obilježjem. Na taj način simulirano je formiranje Bulove mape i njezina elaboracija putem operacija presjeka i unije. Ukoliko se predložena mreža dalje ugradi u veću višerazinsku neuronsku arhitekturu za detekciju rubova ili linija, ona može simulirati prostornu selekciju objekta i implementirati vizualne rutine poput mentalnog praćenja kontura.

U drugom dijelu doktorskog rada razvijena je neuronska mreža za percepciju boje zasnovana na teoriji adaptivne rezonance. Mreža može pomiriti kognitivnu neprobojnost vidne percepcije s neuroznanstvenim spoznajama o povratnim vezama u vidnom korteksu i silaznim utjecajima u vidnoj percepciji. Model objašnjava kako se neuronski signali koje registriraju čunjići u ranim stadijima vidnog procesiranja transformiraju u percepciju boje u kasnijim stadijima. Model također pokazuje da su za ograničavanje utjecaja očekivanja na percepciju boje odgovorni isti mehanizmi koji osiguravaju i stabilnost kodova za percepciju boje. Empirijski opaženi utjecaji očekivanja na percepciju boje objašnjavaju se kao posljedica privremenih nestabilnosti u dinamici mreže, koje dugoročno nestaju. Stoga model pokazuje da očekivanja ne mijenjaju sadržaj svjesne vidne percepcije, barem ne u velikoj mjeri. Općenito, doktorski rad pruža nove uvide u to kako povratne veze posreduju u silaznim utjecajima na vidnu percepciju. Također, predloženi modeli stvorili su brojne hipoteze za daljnja psihofizička i neurofiziološka istraživanja.

Ključne riječi: neuronske mreže, vidna pažnja, silazni utjecaji, kognitivna neprobojnost vida, teorija adaptivne rezonance, percepcija boje

CONTENTS

1

INTRODUCTION

1. INTRODUCTION

1.1. The Role of Feedback Connections in Visual Processing

As early as 1890, William James stated that part of what we perceive comes from our senses, and another, larger part comes from our own mind (Gage & Baars, 2018). Since then, "psychologists and neuroscientists have made remarkable advances in understanding the functional organization of the visual system, uncovering important clues about its perceptual mechanisms and underlying neural codes" (Tong, 2018, p. 1). Visual perception is not just a simple process of registration of the amount of light that impinges upon our eyes. Rather, it involves a set of complex computations that transform noisy and incomplete retinal image into a rich and fairly accurate representation of the external environment (Palmer, 1999). How brain accomplishes this feat is a major challenge for psychology and neuroscience. The traditional view on the computation in the visual system is that it involves a hierarchy of processing stages (Lennie, 1998). Each stage adds a new information to the evolving visual representation by applying the specific algorithm on its input and transmits results to the next stage in the hierarchy. On this view, information flow in the visual system proceeds in a pure feedforward (bottom-up) fashion starting from the retina and ending in the temporal and parietal cortex. However, anatomical and physiological studies revealed many feedback or backward projections from higher to lower levels in the hierarchy (Gilbert & Li, 2013).

Visual processing in the cortex is hierarchically organized and distributed across more than 30 areas (Felleman & Van Essen, 1991; Rockland & Pandya, 1979). Schematic description of visual hierarchy is provided in Figure 1. At the bottom of the hierarchy is the primary visual cortex (V1) with small receptive fields that are tuned to basic visual features such as wavelength, orientation, retinal disparity, and motion direction. As we traverse along the hierarchy to higher-order areas in the inferotemporal and parietal cortex, receptive fields increase in size, complexity, and selectivity (Nassi & Callaway, 2009). Visual processing is characterized by a division of labor between two parallel processing streams: ventral and dorsal (Bell et al., 2014; Milner & Goodale, 2008; Ungerleider & Mishkin, 1982). The ventral stream, also known as the *what* stream, is involved in spatially invariant object recognition. Neurons in the ventral stream are selective for color, orientation, or more complex combinations of colors and line patterns. At the top of the ventral stream is the anterior inferotemporal cortex where neurons exhibit selectivity for complex 2-D or 3-D shapes and objects irrespective of their exact location in space. By contrast, the dorsal stream, also known as the *where* stream, is involved in encoding visual space that supports sensorimotor integration and attention. Neurons in the dorsal stream are selective for motion direction and binocular disparity. At the top of the dorsal stream are areas in the parietal cortex that contribute to the target selection for arm and eye movements, object manipulation and visuospatial attention. Lesions studies further support division between these two streams. The ventral stream lesions impair discrimination of complex shapes and perceptual invariance, while the dorsal stream lesions impair motion perception, speed discrimination, smooth pursuit eye movements, and accurate encoding of visual space. Based on these observations it was proposed that computational goal of the ventral stream is to transform visual inputs into perceptually invariant set of attributes that encode enduring characteristics of the objects, whereas goal of the dorsal stream is to mediate navigation and visual control of skilled actions directed at objects in the visual world (Bell et al., 2014; Milner & Goodale, 2008; Nassi & Callaway, 2009).

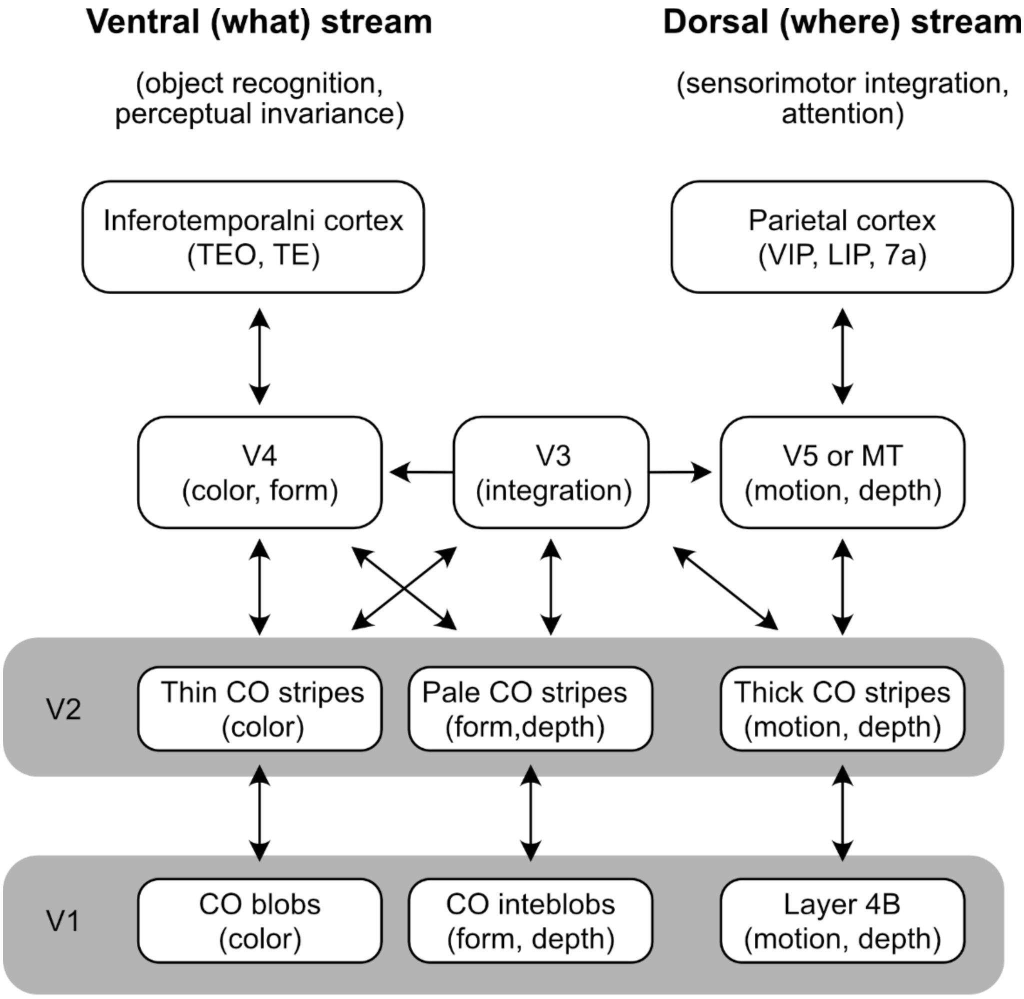

Ventral (what) stream
(object recognition, perceptual invariance)
Dorsal (where) stream
(sensorimotor integration, attention)
Inferotemporalni cortex (TEO, TE)
Parietal cortex (VIP, LIP, 7a)
V4 (color, form)
V3 (integration)
V5 or MT (motion, depth)
V2
Thin CO stripes (color)
Pale CO stripes (form,depth)
Thick CO stripes (motion, depth)
V1
CO blobs (color)
CO inteblobs (form, depth)
Layer 4B (motion, depth)

Figure 1. *Ventral and dorsal streams.*

Prominent feature of the visual hierarchy is a set of feedforward projections from lower- to higher-order areas and more dense set of feedback projections running in an opposite direction, from higher- to lower-order areas in the hierarchy. In a recent review, Briggs (2020) noted that despite many efforts, it was surprisingly difficult to pinpoint exact functional role of feedback projections. For example, feedback synapses from V1 to LGN outnumber feedforward synapses from retina to LGN by almost 10 to 1. Although they are more numerous, feedback synapses are small, and they are located on distal dendrites implying that their postsynaptic currents are attenuated on the path to the soma of the target cell. They also have a low probability of neurotransmitter release and generate small postsynaptic currents. In contrast, feedforward synapses are large, and they are located on dendrites that are closer to the soma of the target cells. In addition, they have a high release probability and generate large postsynaptic currents (Bickford, 2016; Sherman & Guillery, 1998, 2006).

Reviewed properties suggest that feedforward synapses have strong driving impact on the target cell, while the effect of feedback synapses is weak and modulatory. This means that feedforward synapses can drive the cell to a suprathreshold level even in the absence of feedback activity, while feedback synapses cannot do the same in the absence of simultaneous feedforward activity. In accord with the dichotomy between drivers and modulators, it has often been observed that feedback connections cannot alter receptive field structure of the target cell that is derived from their feedforward inputs. Instead, feedback signals provide multiplicative gain modulation to the target cell without altering its tuning preferences (Briggs, 2020).

As shown in the previous paragraphs, it is clear that feedback projections have a distinct role in cortical processing. An important question that remains unanswered is what the exact role of feedback connections in the hierarchical visual processing is and how they contribute to subjective visual perception. Macknik and Martinez-Conde (2009) have suggested that the sole purpose of feedback connections is to maintain visual selective attention on the part of the visual field. Conversely, several authors, inspired by the predictive coding framework, pointed out that feedback projections may communicate prior expectations about upcoming sensory events (Clark, 2013; Hohway, 2013; Friston, 2010). In the thesis, these two hypotheses will be examined from the perspective of computational cognitive neuroscience (Ashby & Hélie, 2011). Through the building of neural network models, consequences of feedback connections on the ongoing visual processing will be explored. The next two sections will review findings on visual selective attention and its computational modeling, respectively. In the Section 1.4., findings on the role of predictions in visual perception from the perspective of predictive coding framework will be reviewed. Focus will be on the issue of cognitive penetrability of vision as a major consequence of the communicating predictions through feedback projections. The Section 1.5 will provide an overview of the adaptive resonance theory that offers an alternative view on feedback connections. This will provide a basis for constructing a neural network model with feedback connections that simulates many properties of visual perception, but that is not cognitively penetrable.

1.2. Visual Selective Attention

Visual attention can be allocated to a circumscribed area in space, to an object irrespective of its exact shape or to an abstract feature value like color or orientation (Nobre & Kastner, 2014). Early studies have shown that location is a unit of visual selection. According to the spotlight metaphor, attention operates like a beam of light that shines over part of the scene, that is, it selects visual information that falls within a single, circular region of space and filters out everything else that falls outside (Posner, 1980). Spotlight of attention has fixed radius that can be moved across visual space to highlight its different parts. Three distinct processes control attentional movement across space: 1) disengaging attention from the currently attended location, 2) moving attention to a new location, and 3) engaging attention at a new location. Each of these processes require time to complete, so they leave distinct signatures (costs and benefits) in the pattern of performance in a detection task depending on how attention is oriented prior to the stimulus presentation. Spotlight metaphor was further augmented by the zoom lens model. It suggests that the radius of the attentional spotlight could be flexibly adjusted, like lens in the camera that can zoom in or out depending on the resolution that observer needs to achieve. In case when higher resolution is required, spotlight shrinks to capture greater details over smaller radius of space. Conversely, when low resolution is sufficient, spotlight widens like in the process of zooming out to pick up less detailed information across larger space (Eriksen & St. James, 1986).

Later studies revealed that the focus of attention does not have to be in the form of a spotlight, but that it can be shaped in more complex ways. For example, it was shown that an object could also serve as a unit of attentional selection, irrespective of its shape or position in space. This is demonstrated in studies showing that focusing attention to one part of the object leads to the processing advantage for other parts of the same object (Chen, 2012). Finally, attention could be allocated to a visual feature like color or direction of motion independent of their spatial location. This form of attention has a global impact on complete visual representation, and not just on specific location (Maunsell & Treue, 2006).

There is also an evidence that object-based attention is linked to spatial representation because observers may select together all locations occupied by the same object (Hollingworth et al., 2012). There are several contrasting explanations of how object-based attention generates this effect. Object-based attention may operate on spatially invariant representation where all object properties are bound together (Vecera & Farah, 1994). Another possibility is that the object-based effects are a mere consequence of attentional prioritization. This means that spatial attention follows sequential path. It is first engaged at the cued location, then it moves to an unattended part of the cued object and finally jumps to the distractor object resulting in the processing advantage of unattended part of the cued object relative to the distractor (Shomstein & Yantis, 2002, 2004). Finally, there is evidence that object-based attention spreads along all locations occupied by the cued object (Richard et al., 2008). In this way, object-based attention selects a grouped array of locations in the retinotopic map (Hollingworth et al., 2012; O'Grady & Müller, 2000; Vatterott & Vecera, 2015).

Roelfsema and Houtkamp (2011) developed an incremental grouping theory (IGT) to explain how grouped array of locations is formed in the visual cortex. On this view, perceptual grouping involves two qualitatively different processes: *base* and *incremental grouping*. Base grouping relies on feedforward connections to accomplish fast segmentation of visual scene in accord with Gestalt grouping cues such as proximity, similarity, closure, and others (Brooks, 2014). In contrast, incremental grouping relies on horizontal and feedback connections to label all connected locations with enhanced firing rate (Roelfsema, 2006). This is a slow, serial process that requires more time to cover larger distances on the map. Incremental grouping is a visual routine that makes explicit those spatial relationships that are not readily available in the base grouping. Examples of such relations are whether two dots lie on the same contour or not, or whether dot lies inside or outside of an enclosed region (Ullman, 1984, 1996). At the neural level, propagation of attentional label is achieved by comparing activity levels of neurons that are subject to attentional modulations (A-units) with another class of neurons that do not receive such modulation and whose response is determined by their classical receptive field (N-units). Such computational scheme enables robust contrast-invariant labeling (Pooresmaeili et al., 2010).

In the same way as object-based attention, feature-based selection may also involve the formation of the spatial map representing all locations occupied by the same feature value. This is illustrated by the pattern matching task where participants must decide whether two spatial patterns are the same or different. Each pattern consists of a 4 × 4 matrix of colored squares. In one condition all squares have unique hue, whereas in the second condition there are four sets of four squares sharing the same hue. Huang and Pashler (2007) found that participants were faster and more accurate in the second relative to the first condition. The same effect also appeared in the symmetry detection task and in the rotation task. This suggests that participants are able to group squares sharing the same hue in the perceptual unit and evaluate it in a single processing step. These observations led Huang and Pashler (2007) to develop a Boolean map theory of visual attention where attentional selection involves division of spatial representation into two mutually exclusive sets. One set represents currently selected locations, and the other set represents all other locations that are not selected. Which locations are selected at any given moment is determined by the top-down guidance. For example, we can choose to attend to all red items in the input. This means that a Boolean map will be formed where all locations occupied by red will be marked as selected (labeled as 1's) and all other locations will be ignored (labeled as 0's).

Based on reviewed findings, it is natural to ask whether there is a single neural mechanism underlying space-, object-, and feature-based attentional selection or they rely on a distinct neural substrate. Given the complexity of visual processing in the cortex and the existence of parallel processing streams, it is likely that attentional selection is distributed across multiple cortical areas. However, there is also an evidence that all three types of selection rely on a common spatial representation. Thus, it is possible that there exists a unique spatial map that can flexibly select location, object, or feature depending on the task demands or top-down guidance. In the thesis, computational requirements that such spatial map needs to satisfy in order to support space- and object- and feature-based selection will be explored. In the next section, current models of visuospatial selection will be reviewed.

1.3. Computational Models of Visual Selective Attention

Computational models of visual attention emphasize bottom-up, image-based control of attentional deployment (Itti & Koch, 2001). They rely on the concept of a saliency map. This is a retinotopic map that encodes how distinctive is each location in the map relative to its background. The saliency map integrates parallel inputs from feature maps that encode visual features such as colors, orientations, directions of motion, and so on. Each feature map detects the presence of its dedicated feature, but it also encodes a feature contrast, that is, how much the detected feature is different from its local neighborhood. For example, strong activity in the feature map for red color is observed when there is a large feature difference in the input when red dot is placed on green background. On the other hand, there would be a weak activity in the feature map if red dot is surrounded by other red dots. The saliency map sends feedforward input to a selection map that chooses location with maximal amplitude in the saliency map. The selection map indicates where the focus of attention is located at a given moment. Saliency computation and selection of the most salient location depends on inhibitory interactions in the neural network. Hence, models of lateral inhibition will be reviewed next.

Lateral inhibition is anatomical arrangement that is found at all stages of visual processing starting from the retina and LGN to the visual cortex (Spillmann, 2014; Spillmann et al., 2015). Lateral inhibition contributes to the formation of the cell's receptive field. For example, in the retina, horizontal and amacrine cells mediate inhibition that form off component of the ON-center OFF-surround and OFF-center ON-surround cells. Cell's name reflects circular excitatory (ON) and inhibitory (OFF) zones that lead to opponent interactions. Similar structures can be found in the LGN and the visual cortex as well. Optimal stimulus for the ON-center OFF surround cell is light spot on dark background, and for the OFF-center ON-surround it is light ring with dark center. Interestingly, the cell is silenced when light covers both ON and OFF components of the receptive field. This suggests that cell encodes stimulus contrast, and not the absolute amount of light that falls on its receptive field. Moreover, the stimulus contrast gets enhanced by lateral inhibition as observed by overshoots and undershoots in the activity profile of retinal cells when they encode luminance step. This property is relevant for understanding many perceptual phenomena including visual illusions (Spillmann, 2014).

In computational modeling, two types of lateral inhibition are often distinguished: feedforward and recurrent or feedback inhibition, as shown in Figure 2 (Levine, 2019, Chapter 4). Feedforward inhibition refers to a feedforward projection of lateral inhibitory collaterals to a next processing stage in the hierarchy. This means that nodes within the same network layer do not interact with each other. Instead, inhibitory effect is expressed in the next processing stage. Feedforward inhibition exhibits simple dynamics because target nodes always converge to a fixed point. Feedforward inhibition is employed in modeling early visual processing including spatial, brightness, and motion perception (Carpenter & Grossberg, 1991; Grossberg, 1988). In modelling visual attention, feedforward inhibition is used to compute saliency or feature contrast within maps encoding visual features (Itti & Koch, 2001). For example, suppose that an input consists of red dot that is surrounded by many green dots. In this case, red dot is more salient, that is, it pop-ups on the background of green dots. To account for this effect, it is sufficient to assume independent feedforward lateral inhibition in the map that encodes red color and in the map that encodes green color. Because there is just one red dot, total amount of lateral inhibition in the red map will be small resulting in the strong activity at the location of red dot. In contrast, there are many green dots that will produce strong lateral inhibition among themselves and consequently a low level of activity at the locations corresponding to them. As a consequence, red dot will stand out by higher level of activity in the saliency map relative to green dots and it will be selected first.

Another form of lateral inhibition is recurrent inhibition (Levine, 2019, Chapter 4). It refers to mutual lateral interactions within the same layer of nodes. The network with recurrent inhibition typically consists of an array of excitatory nodes which are reciprocally connected with the small population of inhibitory interneurons. In this case, more complex dynamics is possible including oscillations and chaos (Ermentrout, 1992). The type of dynamics that the network will display is determined by its parameters, that is, synaptic weights that control the strength of self-excitation and lateral inhibition. In certain parameter ranges, the network with recurrent inhibition will settle to a fixed point where a single node remains maximally active, while all other nodes are inhibited to zero (Rutishauser & Douglas, 2009; Rutishauser et al., 2011). The winner is a node that received the strongest input. This is an extreme form of competition known as a *winner-takes-all (WTA) behavior*. The WTA network makes a binary decision among inputs it receives. They are employed in modeling decision making, object recognition, and visual attention (Carpenter & Grossberg, 1991; Grossberg, 1988). In modeling visual selective attention, the WTA network is used to select the most salient location to which attention should be directed (Itti & Koch, 2001).

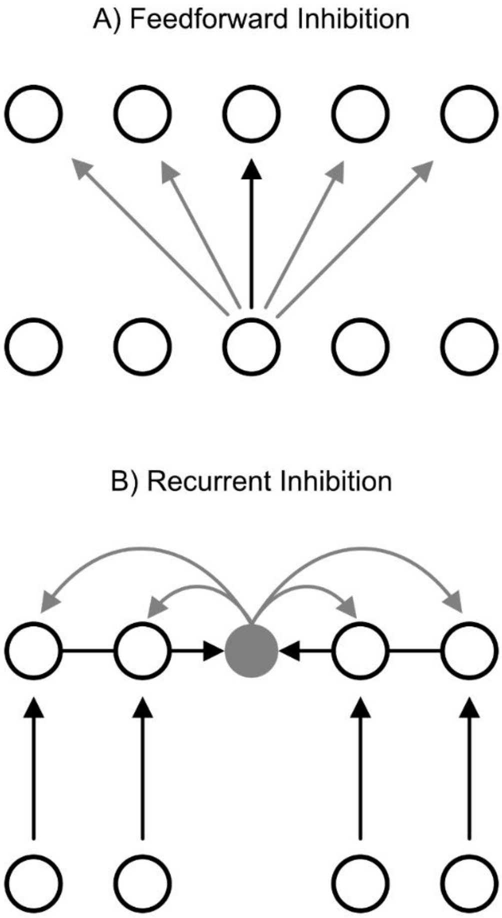


Figure 2. *(A) Feedforward inhibition and (B) recurrent inhibition. Black arrows depict excitatory connections, whereas gray arrows depict inhibitory connections. Self-excitation is not shown.*

Important problem faced by the WTA network is that it quickly loses its capacity to represent winners as the number of winning nodes increase (Haarmann & Usher, 2001; Usher & Cohen, 1999). This feature is undesirable if multiple winners should simultaneously remain active. Such situation arises in modeling object- and feature-based selection where the WTA network must select all locations occupied by the object or all locations occupied by the same feature value. The reason for the inability of the standard WTA network to represent multiple winners is that the total amount of lateral inhibition among nodes increases proportionally to the number of active nodes. Consequence of raising inhibition is that all nodes, even those receiving strong input, will be eventually silenced to zero making the WTA network incapable of representing winning object or feature. This problem cannot be solved simply by reducing the strength of lateral inhibition because weak inhibition will allow non-winning nodes to remain active also. In addition, it is not possible to know in advance what is the size of the winning object or how many locations the winning feature value contains to precisely adjust the inhibitory weights. Instead, more principled way is required to deal with this problem. In particular, it is important to find a way to control excessive inhibition. The problem of multiple winners will be analyzed in greater detail in this thesis. The thesis will offer a solution that will greatly expand the capability of the WTA network to flexibly select winners based on the task demands. In particular, two biophysically plausible neural components that can solve the problem of simultaneously representing arbitrary large number of winning nodes without sacrificing the network ability to make selection will be proposed.

1.4. The Effect of Expectations on Visual Perception

The second hypothesis that is going to be examined in the thesis is that feedback projections transmit expectations about upcoming sensory events from higher to lower stages in the visual hierarchy. This hypothesis arises from the predictive coding framework suggesting that visual perception is inferential process because the brain is designed to perform statistical (Bayesian) inference. Predictions encode prior knowledge about the environment and help to disambiguate noisy retinal signals. Beliefs about the current state of the environment are updated under observed sensory evidence (Clark, 2013; Hohwy, 2013). For example, light projected on the retina loses information about the distance from which it is emitted. Thus, the brain needs to use prior knowledge about the retinal disparity of image elements to recover information about the depth from the observer. Such reconstruction is of critical importance for survival because it allows individual to successfully navigate through the 3-D space. Sometimes, the generative model does not fit the sensory data very well and the visual system generates prediction error that should be minimized by revising the generative model. A principled way of how to optimize the model in light of new data is prescribed by the Bayes rule, which specify how to compute posterior probability of a given hypothesis based on its prior probability and observed evidence. According to a closely related free energy principle, prediction error is undesirable because it signals that the individual lost homeostatic balance with the environment. Free energy formulation also helps to explain how to optimize generative models in a hierarchical architecture where predictions and prediction errors flow across multiple processing stages (Friston, 2008, 2010).

Predictive coding framework assigns specific functional roles to feedforward and feedback projections (Bastos et al., 2015; Keller & Mrsic-Flogel, 2018; Shipp, 2016). Feedback projections communicate predictions or expectations about incoming sensory stimulation to the next lower area, whereas feedforward projections compute prediction error and propagate it to a next higher level in the visual hierarchy. On this view, feedforward activity should be strong when visual system is faced with unpredictable stimulus because its activity is proportional to the deviation of sensory signals from the prediction. In contrast, predictable stimulus should generate only weak feedforward activity because there is no deviation, and feedback projections successfully explained away what caused it. Hence, there is nothing new to communicate to a higher level in the hierarchy. Additional advantage of this processing scheme is that, with the predictable stimuli, feedforward projections do not generate spike trains and consequently save the costly energy (Aitchison & Lengyel, 2017). Computational implementations of predictive coding have been successful in modelling some of the response properties of cells in visual cortex such as end-stopping and biased competition (Rao & Ballard, 1999; Spratling, 2008, 2010).

Several authors have pointed out that predictive coding implies *cognitive penetrability of vision (CPV)* suggesting that generative models directly modulate and shape visual perception (Lupyan, 2015a, 2017a; Newen & Vetter, 2017; O'Callaghan et al., 2017; Vetter & Newen, 2014). For example, Lupyan (2015a, p. 547) explicitly stated "that expectations, knowledge, and task demands can shape perception at multiple levels, leaving no part untouched." This follows from the global minimization of prediction error that constrains bottom-up activity by top-down knowledge. Moreover, he claimed that cognitive malleability of perception implies that our belief in reliability of perception is misplaced. Visual perception is as good at picking up information in the environment as its generative models allows it to be (Lupyan, 2017b). Along the same lines, O'Callaghan et al. (2017) reviewed evidence suggesting that predictions are derived from a rich source of cognitive, affective, and contextual associations that directly influence visual perception because they carry important information that is missing from raw sensory stimulation. On this view, visual perception requires delicate balance between bottom-up and top-down signals. Disruption of this balance, as observed in some neuropsychiatric disorders, leads to visual hallucinations, which provides further support for the importance of top-down processes in visual perception. O'Callaghan et al. (2017, p. 63) concluded that "…predictive penetration, be it cognitive, social or emotional, should be considered a fundamental framework that supports visual perception". Many behavioral and neuroimaging studies of top-down effects on visual perception support this view by showing that knowledge, beliefs, actions, and emotions do indeed alter the content of visual experience (reviewed in O'Callaghan et al., 2017). For example, information about which of the two oriented gratings were more likely to occur in a display heightens visual sensitivity to a cued orientation during fine orientation discrimination (Cheadle et al., 2015). Functional neuroimaging showed modulations in V1 activity that was correlated with top-down predictions (Edwards et al., 2017).

However, behavioral studies supporting CPV have been criticized because of the lack of methodological rigor (Firestone & Scholl, 2016). Important obstacle in testing CPV is that behavioral response does not need to directly reflect visual perception because it is simultaneously influenced by several extra-visual sources. For example, wearing heavy backpacks lead participants to overestimate the steepness of the hill suggesting that kinesthetic information provided by the muscles modulates perception of the slant (Bhalla & Proffitt, 1999). However, it may simply reflect experimental demands because it was obvious to participants what is the hypothesis that has been tested. When the relationship between backpack and perceptual task is made less obvious, the effect of backpack disappeared (Durgin et al., 2009). It seems that there is a general tendency in studies reporting CPV to confuse visual perception with judgment or decision making (Firestone & Scholl, 2017).

The same problem of a failure to distinguish between perception and decision making arises even in studies examining the effect of prior knowledge on such basic perceptual attribute such as color. Several studies reported on so called memory color effect where color adjustments of a grey object are biased toward complement of typical color in which the object appears (Hansen et al., 2006; Olkkonen et al., 2008; Witzel et al., 2011). For example, participants adjusted grey banana to the bluish hue to offset perceived yellow that is retrieved from memory. However, it is possible that memory color effect arises from indeterminacy of instructions to participants and their conscious decision to offset the adjusted color from neutral gray just to be safe (Zeimbekis, 2013). Valenti and Firestone (2019) eliminated such response bias by asking participants to choose an odd item from a set of three alternatives (gray banana, bluish banana, and gray object without diagnostic color). In this condition, participants should choose gray banana as an odd item because it should appear yellowish, while blue banana should appear neutral gray and, consequently, indistinguishable in the hue from gray object without diagnostic color. However, participants predominantly choose bluish banana as an odd item suggesting that they were not influenced by the retrieval of color from memory.

Another concern is that behavioral and neuroimaging studies were often conducted with a small number of participants, leading to insufficient statistical power to detect the actual effect. Consequently, researchers sometimes resort to a search for patterns in the data instead of testing the hypothesis. This is not the way how hypothesis testing should be done (Francis, 2019). Such strategy may lead to false positive findings regarding CPV. For example, in a recent attempt to find perceptual signature of a prediction error computation, Staadt et al. (2020)

reported on the new illusion that observers experience when they briefly viewed at the superposition of two orthogonally oriented gratings. After abruptly removing one of the displayed orientations, instead of strictly seeing the remaining orientation, observers reported an illusory percept of the arithmetic difference between previous and actual orientation. However, the effect appeared in only 7 out of 15 participants. Moreover, even those participants who reported seeing the illusion noticed it only occasionally.

From the theoretical point of view, it is important to distinguish between early and late vision (Marr, 1982; Pylyshyn, 1999). Early vision refers to a formation of surface representation in depth. Early vision is associated with neural processing in lower portion of visual hierarchy including areas V1, V2, V3, V4, and MT. Conversely, late vision refers to object recognition that is associated with top levels of visual hierarchy such as the inferotemporal cortex. While there is a general agreement that late vision is cognitively penetrable, there is a controversy whether the same conclusion could be extended to early vision (Raftopoulos, 2019; Raftopoluos & Zeimbekis, 2015). According to Pylyshyn (1999), early vision is informationally encapsulated module with fixed, innate architecture. It is based on specific set of principles that are fundamentally different from those governing central cognitive functions such as thinking or reasoning. Now, the question arises of how early vision maintains its independence and separation from top-down influences in light of overwhelming evidence for feedback communication in the visual cortex. The answer provides a different theoretical framework that will be described in the next section.

1.5. Adaptive Resonance Theory as an Alternative to Predictive Coding

To provide a computational argument that feedback projections in the visual cortex do not imply cognitive penetration of visual perception, adaptive resonance theory (ART) is adopted as the guiding theoretical framework in the thesis (Grossberg, 2013, 2017). The ART is primarily designed to address the problem of stability of learning and memory in a dynamic environment. Many neural network algorithms suffer from catastrophic forgetting because learning new input patterns erases old memories. According to Grossberg (2013), the solution to the problem of catastrophic forgetting is to compare sensory (bottom-up) data with learned (top-down) expectations, which are transmitted via feedback projections. If the input pattern is matched with one of the previously learned codes (categories), it is recognized as a familiar pattern. Familiarity with the pattern is signaled by the resonance or mutual excitatory reinforcement between the sensory pattern and top-down activation. Grossberg (2017) argued that the resonant state in the network corresponds to our conscious visual experience. On the other hand, if there is a mismatch between the input pattern and the learned expectation, a special novelty detection system is activated that resets the currently activated expectation and thus erases its trace from the network activity.

As can be seen from previous description, ART handles the match or mismatch between the bottom-up input and top-down expectations in a different way relative to predictive coding. In ART, the match is reinforced by mutual excitation between feedforward and feedback projections leading to a resonant or conscious state. In contrast, the mismatch leads to activation of the reset mechanism that removes erroneous prediction from ongoing processing in the ART circuit. Thus, top-down predictions cannot influence visual perception in an arbitrary way as suggested by Lupyan (2015a) and O'Callaghan et al. (2017). In ART, the neural processing is driven more by bottom-up input. This property will be used to argue that the ART is a predictive system that is not cognitively penetrable. In the thesis, a specific implementation of the ART circuit will be developed and used to simulate some of the effects taken as evidence for CPV.

1.6. Research Aim, Objectives, and Problems

The general aim of the doctoral thesis is to develop new neural network models that will clarify how top-down effects such as visual selective attention and expectations affect visual perception. The models will be tested and evaluated through a set of computer simulations.

Specific objectives of the doctoral thesis are:

1. To develop a recurrent competitive neural network that is capable of flexibly selecting a location, object, or specific feature value depending on the top-down guidance. It will be examined how to extend the standard model of WTA network to enable simultaneous selection of multiple winners without degrading its representational capacity. The proposed model should unite the Boolean map theory of visual attention (Huang & Pashler, 2007) with the incremental grouping theory (Roelfsema & Houtkamp, 2011) to provide a comprehensive description of how visuospatial selection is implemented in the visual cortex.
2. To develop a neural network model based on the adaptive resonance theory that can simulate behavioral findings on the top-down influences on visual perception. The model should reconcile cognitive impenetrability of visual perception with anatomical and functional data on the existence of massive feedback projections in the visual cortex.

In this way, the thesis will explore to what degree ART is a viable alternative to predictive coding framework in understanding the role of feedback projections in cortical information processing.

1.7. Research Methodology

Computational cognitive neuroscience is an interdisciplinary scientific domain that integrates knowledge from psychology, neurobiology, mathematics, and computer science to study the relationship between mind and brain. Theoretical models of neural networks provide a unique perspective on addressing major challenge: How complex brain interactions generate intelligent behavior? They enable rigorous quantitative analysis of how neurons and their synaptic connections give rise to cognitive functions such as visual perception and attention (Ashby & Hélie, 2011; Levine, 2019; O'Reilly & Munakata, 2000).

Computer simulations of neural network models are employed as the main research method. The method consists of the following steps:

1. Specification of the mathematical model of a neural network designed to solve specific perceptual or cognitive task,
2. Formal analysis of the model and its properties (stability analysis, the existence of fixed points),
3. Translation of the model into the simulation software and its comprehensive testing on computer,
4. Comparison of the model output with behavioral findings on humans and/or primates.

The proposed neural networks will be implemented in the programming language for scientific calculations MATLAB. As an input to the neural networks, stimuli similar to those used in published psychophysical or neurophysiological experiments will be employed to facilitate comparison with empirical studies. Network models will be described by the system of non-linear differential equations that specify how neural activity evolves over time. Computer simulations will involve numerical integration of differential equations using Euler's method with fixed time steps.

1.7.1. WTA

To address the first objective, a new model of lateral inhibition that simultaneously selects multiple locations in a 2-D spatial map representing visual space will be developed. Previous models of visuospatial selection rely on extreme form of competition known as a WTA network, described in Section 1.3. The WTA network consists of an array of excitatory nodes which are reciprocally connected with the small population of inhibitory interneurons (Rutishauser & Douglas, 2009; Rutishauser et al., 2011). This anatomical arrangement leads to lateral inhibition between excitatory nodes and consequently to the selection of single node receiving the maximal input. Such behavior is consistent with space-based attention, but it is inadequate to handle object- or feature-based attention because these forms of attention require simultaneous selection of arbitrary many locations in the spatial map. Excessive amount of recurrent inhibition in the WTA network has been identified as a major obstacle to select multiple winners (Haarmann & Usher, 2001; Usher & Cohen, 1999).

In the thesis, it is proposed that the WTA network can correct this problem by engaging additional mechanisms such as synaptic (Abbott & Regehr, 2004) and dendritic computation (Häusser & Mel, 2003; London & Häusser, 2005). Recurrent excitation among WTA nodes should be mediated via dendrites whose output nonlinearity will prevent instability or unbounded growth of neural activity caused by the positive feedback loop. Furthermore, the inhibitory interneuron should compute function maximum instead of the sum over its input from the excitatory nodes. This is achieved by presynaptic inhibition that acts upon synaptic transmission and regulates the amount of transmitter release in an activity-dependent manner. The proposed computational elements increase the stability of the WTA network, but at the same time increase its flexibility in shaping attentional selection according to the task demands.

1.7.2. ART

To address the second objective, a real-time implementation of the adaptive resonance theory (described in Section 1.5.) will be developed to explain how color opponent responses in the early stages of visual processing are transformed into color categories in the later stages. It will be demonstrated how feedback projections contribute to stable recognition and conscious experience of colors. However, it will also be shown that the same mechanisms that enable stability of learning also prevent expectations to influence perception directly. Instead, the observed empirical effects of expectations and prior knowledge on color perception will be attributed to temporary instability in the network dynamics while seeking the best match between the bottom-up input and top-down expectations. In this way, it will be demonstrated that the adaptive resonance theory (Grossberg, 2013, 2017) offers an attractive alternative theoretical framework that explains various examples of top-down effects on perception in a fundamentally different way compared to the currently dominant predictive coding model (Clark, 2013; Hohwy, 2013).

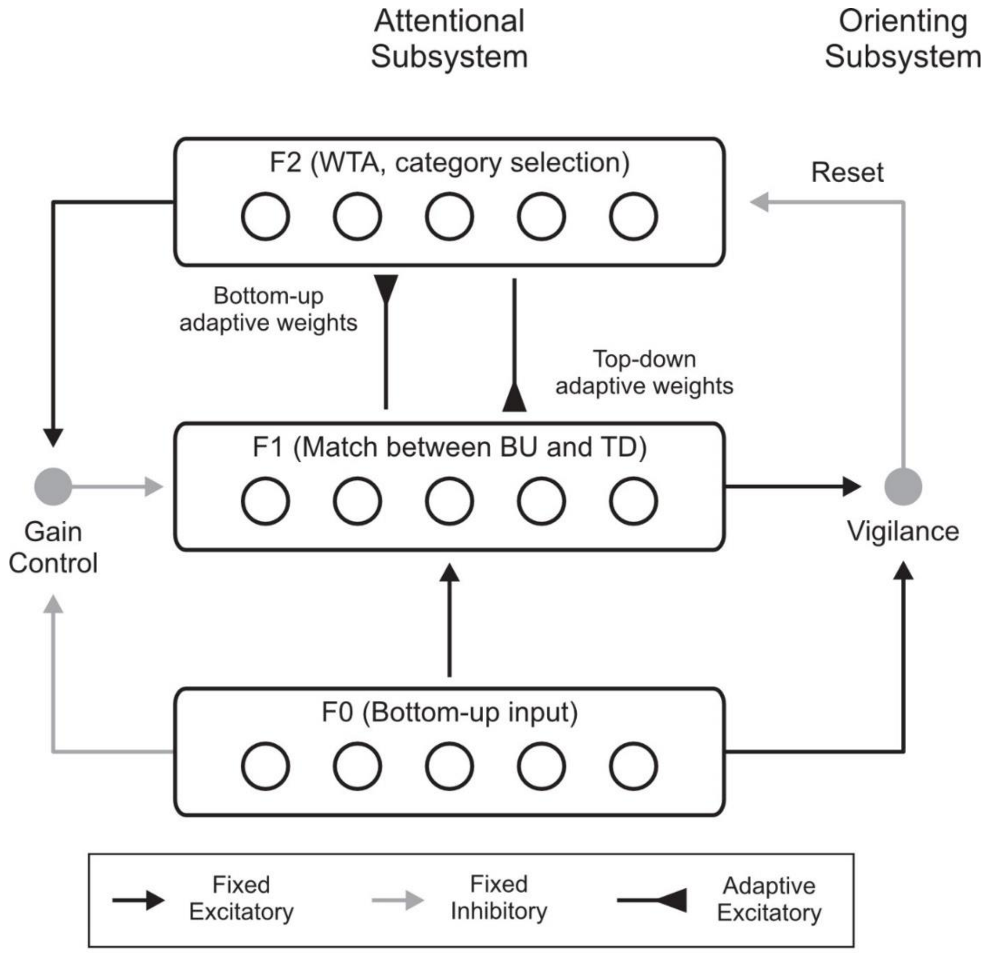


Figure 3. *Adaptive resonance theory (ART) circuit.*

ART is a three-layer architecture with two auxiliary mechanisms for controlling network activity: a gain control mechanism and a reset mechanism (Figure 3). Layers are denoted as F_0, F_1 and F_2 (Carpenter & Grossberg, 2003). The F_0 layer is an input layer that registers the pattern of sensory stimulation. Next, the F_1 layer reads-out the sensory pattern from F_0 and combines it with the top-down expectations arriving from the F_2 layer. The activation flows from F_1 to F_2 and passes through a filter of adaptive weights. Finally, the F_2 layer is a winner-takes-all network that represents the category or concept that best matches the sensory input. The Gain

Control mechanism enables to distinguish between sensory stimulation and internal activation. Only sensory stimulation is allowed to reach the supra-threshold activation and ignite the resonance between F_1 and F_2 layer. Finally, the Orienting Subsystem monitors for the difference between the bottom-up activation from F_0 and the top-down activations from F_2. It produces a reset signal when the difference is larger than a certain prescribed value controlled by the vigilance parameter. Reset signal shuts-off the currently active F_2 node and initiates a search for another F_2 node that will provide a better match with the sensory pattern.

2

ELABORATION

2. ELABORATION

The doctoral thesis is written in the format of a compilation thesis and consists of a compendium of four scientific papers published in indexed journals. This section provides comprehensive overview of research questions, methodology, results, and conclusions presented in these papers to document the coherence of the thesis.

2.1. An Overview of Studies

The study of Marić and Domijan (2018a) and Marić and Domijan (2019) addressed the first problem of the thesis, i.e., to examine whether it is possible to design a new recurrent competitive neural network model of visual selective attention that can select a location, object, or specific feature value depending on the task demands. The standard approach to model visual selective attention is to employ recurrent competitive network known as the WTA network. It implements extreme form of competition where only a single node, receiving maximal input, survives competition and remains active in the network, while all other nodes are silenced to zero. Such model is not adequate to simulate properties of object- and feature-based attention because they require simultaneous selection of multiple units. To address this issue, standard WTA network is extended by including two new computational elements: dynamics regulation of synaptic transmission and dendritic nonlinearity. Computer simulations revealed that the proposed model successfully selects all locations occupied by the chosen feature as suggested by Boolean map theory of visual attention (Huang & Pashler, 2007). In addition, the model exhibits object-based selection when one location of object is cued by top-down signals. In this case, the enhanced activity spreads from cued location to all connected locations, thus selecting all locations occupied by the cued object. In this way, the proposed model implements contour tracing, an important visual routine that enables visual system to segment relevant from irrelevant objects in a scene populated with many objects (Roelfsema & Houtkamp, 2011).

The review paper of Marić and Domijan (2018b) and the study of Marić and Domijan (2020) provided an answer to the second problem of the thesis, i.e., to examine whether it is possible to design a neural network model that can reconcile cognitive impenetrability of visual perception with neuroscientific findings on feedback connections in the visual cortex. The answer to this question is positive because there is a well-studied neural architecture known as the adaptive resonance theory that is capable of stable learning and recognition of input patterns. Here, a specific implementation of the ART circuit is designed to simulate color perception, thus it is named the color ART circuit. Computer simulations of the color ART circuit showed that it can select the hue category that is the most consistent with its cone-opponent input. Top-down signals arising from the object-recognition network may temporarily disrupt processing in the color ART circuit and bias its response toward the expected hue. However, top-down connections cannot have long-lasting effect because reset mechanism that assures stability of learning clears them off. Thus, experimentally observed effects of top-down influences on color perception are ascribed to the uncertainty that arises in the short period of time after the stimulus presentation when the color ART circuit has not yet found the hue category that best matches the input pattern.

Together, these models will show that visual perception is not cognitively penetrable although its neural substrate involves the coordinate interplay between feedforward and feedback connections.

2.1.1. A Model of Feature-Based Spatial Selection

Marić and Domijan (2018a) developed a new model of the WTA network that can simultaneously select multiple spatial locations based on a shared feature value. The model is named the feature-based WTA (F-WTA) network because the unit of selection is not a point in space or object, but rather an abstract feature value that is set by top-down signals. This study demonstrates how the F-WTA network implements the central proposal of Boolean map theory of visual attention (see Appendix A for the full paper and Appendix E for supplemental materials containing the MATLAB code to reproduce all results).

2.1.1.1. Introduction

According to Huang and Pashler’s (2007) Boolean map theory of visual attention, feature-based attention involves the selection of all locations occupied by the same feature value (e.g., red) per dimension (e.g., color). When forming a Boolean map, all spatial locations occupied by the chosen feature value are selected, and all others are ignored. In other words, the Boolean map controls spatial selection and access to consciousness. After the formation of a Boolean map, it is possible to operate on its output by applying the set operations of intersection and union. A conjunction search task in which two feature dimensions (e.g., color and orientation) should be combined to find the target object (e.g., red horizontal bar) is an example of the intersection of two Boolean maps. First, a Boolean map is formed by top-down cueing of the red items irrespective of their orientations. Second, observer cues the horizontal items among these selected red items, so the resulting Boolean map will represent the intersection of red and horizontal items. Similarly, the union between two Boolean maps is achieved by top-down cueing the red items in the first step and then cueing the horizontal items in the second step. But the locations selected in the first step are simultaneously maintained in memory so that the resulting Boolean map can simultaneously represent locations of all red and all horizontal items.

Computational models of visual selective attention rely on the WTA network to select only the most salient location in the input image. The WTA network that is capable of computing with Boolean maps should simultaneously select arbitrarily many locations that share a common feature value without degrading the representation of winners. Likewise, the WTA network should be capable of state-dependent computation in which new inputs are combined with the current memory state. The standard WTA network cannot achieve these computational goals because it suffers from the capacity limitations and it is not capable of state-dependent computation. To deal with these problems, a new WTA network is being developed that provides a neural implementation of the Boolean map theory of attention.

The aim of this study was to provide an explanation of how a Boolean map may be formed in a recurrent competitive network that can implement feature-based winner-takes-all (F-WTA) selection. This is achieved by extending the WTA model based on linear-threshold units (Hahnloser, 1998; Hahnloser et al., 2003; Rutishauser & Douglas, 2009). This WTA model consists of a single inhibitory unit reciprocally connected to a group of excitatory units. Here, we elaborated on this basic WTA design by introducing two processing components: dendritic non-linearity and retrograde inhibitory signaling. These biophysical mechanisms are plausible candidates for computation in real neural networks. The dendritic tree is modeled as a separate electrical compartment with its own non-linear output that is supplied to the node’s body. This component is introduced to prevent excessive excitation that arises from self-recurrent and nearest-neighbor collaterals (Branco & Häusser, 2010; London & Häusser, 2005; Häusser & Mel, 2003; Mel, 2016). On the other hand, retrograde signaling crates a feedback loop that dynamically regulates the amount of neurotransmitter released from presynaptic terminals. This component is introduced to isolate winning nodes from mutual inhibition (Alger, 2002; Regehr et al., 2009; Tao & Poo, 2001; Zilberter et al., 2005). In this way, the amount of inhibition in the network is significantly reduced because the inhibitory interneuron computes the maximum instead of the sum of its recurrent excitatory inputs. Consequently, winning excitatory nodes release their retrograde signals and block inhibition from the inhibitory interneuron. Thus, irrespective of the number of winning nodes or their spatial arrangement, arbitrarily many winners can be simultaneously selected.

We followed the model of Hamker (2004) to describe and illustrate cortical computations underlying attentional guidance by top-down feature-based cues. The proposed F-WTA circuit is situated in a larger neural architecture consisted of the model of cortical area V4, the inferotemporal cortex (IT), the posterior parietal cortex (PPC), and the frontal eye fields (FEF). Top-down signals that provide feature cues originate in IT that contains a spatially invariant representation of relevant visual features. IT sends feature-specific feedback projections to V4 where topographically organized feature maps representing each feature value are located. We hypothesized that the F-WTA network resides in PPC where it receives summed input over all feature maps from V4. Top-down guidance is implemented by a temporary increase in activity amplitude in one of the V4 feature maps. Top-down signals to the feature map are modeled as a multiplicative gain of neural activity, which is consistent with neurophysiological findings (Martinez-Trujillo & Treue, 2004; Maunsell & Treue, 2006; Treue & Martinez-Trujillo, 1999).

A set of computer simulations was performed to illustrate the model's behavior. The input was delivered as a vector of 200 excitatory units and one inhibitory unit. Differential Equations were solved numerically using MATLAB's *ode15s* solver. The simulations were run for 250 arbitrary time steps. First, a simulation of the formation of a single Boolean map was performed to demonstrate how it arises in the F-WTA network in response to the presentation of the color cue. Next, a simulation of the intersection and union of two Boolean maps was performed to demonstrate how the model achieves these set operations. Finally, a simulation of bottom-up spatial selection was performed to show that, in the absence of top-down guidance, the network will select the most salient locations based on the bottom-up input.

2.1.1.2. Simulation of the Formation of a Single Boolean Map

In the simulation of the formation of a single Boolean map, the input consisted of red and green items of equal sizes intermixed in space on the black background. Top-down gain was applied on red items at t = 50, which meant that red color was attended. Consistent with findings that feature-based attention spreads across the whole visual field (Saenz et al., 2002, 2003; Serences & Boynton, 2007), top-down gain was also applied to the empty space between items. The duration of the top-down cue was 50 simulated time steps. Subsequently, at $t = 150$, green color was cued in the same way.

At the beginning of the simulation, the F-WTA network simply selected all presented items together, irrespective of their color. This is evident in the first 50 ms of the simulation. In the next step, when red color was cued by applying top-down signals to the corresponding feature map (at $t = 50$ ms), the network created a Boolean map by selecting the spatial pattern associated with red color. Due to self-excitation, the network maintained locations of the cued feature value in working memory even after the cue was removed. When the observer decides to switch attention to another feature value (at $t = 150$ ms), the network selects locations of the new feature value (green) and suppresses locations associated with the previously cued value (red) without requiring an external reset. Namely, the network is sensitive to input changes even though it also exhibits activity persistence. Importantly, the activity level at the selected locations is invariant with respect to the number of active nodes. This is a consequence of retrograde inhibitory signaling in the recurrent pathways between excitatory and inhibitory node.

Additionally, in this simulation it was shown that the F-WTA network can support space- and object-based attention alongside feature-based attention. When the spatial cue was applied to a single location in one of the feature maps, the network responded by selecting only this location. Neighboring nodes were not selected because they received weaker input relative to the cued node. Interestingly, when the spatial cue was removed, the network activity started to propagate from the cued node towards the boundary of the whole object, e.g., all locations that are connected to it. Therefore, the F-WTA network exhibits object-based selection. This finding is consistent with neurophysiological studies that have found spreading of enhanced activity along the shape of the object (reviewed in Roelfsema, 2006; and Roelfsema & Houtkamp, 2011). Moreover, this simulation shows that spatial attention can be easily oriented toward a new location in a single jump without the need for attentional pointers that move attention across the map as in the model of Hahnloser et al. (1999).

2.1.1.3. Simulation of the Intersection and Union of Two Boolean Maps

In the simulation of the intersection of two Boolean maps the visual input consisted of red and green horizontal, and red and green vertical bars. In the first step, the F-WTA network was cued to select red bars, irrespective of their orientation. In the second step, network was cued to select horizontal bars, irrespective of their color. Since the green vertical bars were already suppressed because they received a weak top-down signal, the network output was the selection of a subset of the red bars that are also horizontal. In other words, the network activity converged to the intersection between a set of red and horizontal bars.

In the simulation of the union of two Boolean maps, the visual input consisted of two non-overlapping components: colored squares that activate only feature maps for color and achromatic horizontal and vertical bars that activate only feature maps for orientation. Red colored items occupied locations between 1 and 100, and oriented bars between 101 and 200, which closely resembles Huang and Pashler's (2007) stimulus. Red color was cued in the interval [50, 100], and horizontal items were cued in the interval [110, 160]. This led to the formation of the union of red and horizontal items, that is, to the simultaneous selection of all red and all horizontal items presented in the input. However, we have demonstrated that if top-down cues do not follow each other closely, the second cue overrides the network activity remained from the first cue. Taken together, the results showed that the union of two Boolean maps was possible to achieve only when two top-down cues partially overlap in time or when the second cue closely follows the withdrawal of the first cue.

2.1.1.4. Simulation of Bottom-Up Spatial Selection

In the simulation of bottom-up spatial selection, different input magnitudes were arbitrarily assigned to different items. As illustrated in the paper in Appendix A, without top-down guidance, the F-WTA selected the most salient object. This ability depends on the difference in the input magnitudes. The simulation illustrates that the difference between the two most active nodes should be sufficiently large to separate them. Otherwise, the two most salient items would be chosen. The precision of saliency detection depends on the threshold for the activation of the synaptic receptors on the inhibitory interneuron. If the values were smaller than $T_y = 0.1$, the network would improve its precision and could separate two objects but would lose the ability to form a union of two Boolean maps.

In addition, this simulation showed that the network is sensitive to the sudden appearance of a new object in the scene, suggesting that it can be guided by bottom-up feature cues in the absence of top-down cues (Theeuwes, 2013). Such bottom-up guidance may result from strong inputs arriving from the transient channel. Transient channel can override the network's current memory state and enable the network to reorient attention and select a new object that suddenly appears in the visual scene (Kulikowski, 1973). When the abrupt onset produces only a weak transient signal, the F-WTA network activity stays on the previously attended item, which is consistent with the behavioral findings that the abrupt onset can be ignored in some conditions (Theeuwes, 2013).

2.1.1.5. Discussion

When comparing the F-WTA network with previous models of WTA behavior, we have found that F-WTA has several advantages. First, F-WTA successfully integrates information across space and time to form the intersection or the union of two Boolean maps. Second, previous WTA models require external inhibition of the current winner, or the introduction of a dynamic threshold or habituation to allow attentional focus to move between locations (Horn & Usher, 1990; Itti & Koch, 2000, 2001; Kaski & Kohonen, 1994). In the F-WTA model, there is no need for such external inhibition or habituation. Instead, dendritic and synaptic nonlinearities ensure that the network dynamics is always sensitive to the top-down or bottom-up guidance. Third, F-WTA employs dynamic quenching threshold (QT) like in the shunting competitive model (Grossberg, 1973) where all nodes whose activity is above QT are enhanced, and others below QT are suppressed. Fourth, F-WTA can select and store arbitrarily many locations in memory, which is consistent with the behavioral findings that our ability to select multiple objects is not fixed (Alvarez & Franconeri, 2007; Davis et al., 2000; Davis et al., 2001; Liverence & Franconeri, 2015; Scimeca & Franconeri, 2015). Other WTA networks have limited capacity to represent multiple winners because they lack a mechanism for controlling the inhibition between the winning nodes (Usher & Cohen, 1999).

On the other hand, limitation of the proposed model is that it does not implement all aspects of the theory proposed by Huang and Pashler (2007). Specifically, it does not explain why attention is limited to only one feature value per dimension or how the observer sequentially chooses one feature value after another. Also, in all reported simulations, items were segregated in space, unlike Huang and Pashler (2007) who employed a matrix of connected squares. Finally, the input to the network does not follow the distance-dependent activity profile that is usually observed in the visual cortex. However, this is not a critical issue for the model's performance because the precision of selection is controlled by the thresholds for presynaptic terminal activation.

In summary, this study demonstrates how the feature-based WTA network achieves spatial selection of all locations that are occupied by the same feature value without suffering from capacity limitations. The network responds to the top-down cue by storing spatial pattern in short-term memory. The spatial pattern corresponds to the cued feature value, while non-cued feature values are suppressed. In this way, we showed how the Boolean map is formed. Additionally, computer simulations showed that it is possible to create more complex spatial representations that involve the intersection or union of two or more Boolean maps. In this way, the F-WTA network goes beyond the capabilities of previous models of the competitive neural network that cannot integrate information across space and time. This study suggests that dendritic nonlinearity and retrograde signaling are biophysically plausible mechanisms that are essential for the model success. To conclude, reported simulations suggest that the proposed model of the F-WTA network successfully implemented the Boolean map theory of visual attention and addressed the first part of the first problem of the thesis.

2.1.2. A Model of Object-Based Spatial Selection (Mental Contour Tracing)

The aim of the Marić and Domijan's (2019) study was to examine whether the same model of the feature-based WTA network (proposed in Marić and Domijan, 2018a) supports the implementation of a complex cognitive operation such as mental contour tracing. This study demonstrates how to embed the F-WTA network into a larger neural architecture incorporating multi-scale contour and L-junction detection networks (see Appendix B for the full paper and Appendix E for supplemental materials containing the MATLAB code to reproduce all results).

2.1.2.1. Introduction

Mental contour tracing is an example of a visual routine (Ullman, 1984, 1996) or incremental grouping process (Roelfsema & Houtkamp, 2011) that has been extensively studied at both psychological and neural levels. It is engaged when the observer attempts to determine whether two image regions are connected or not. The detection of connectedness is important because connected image parts are likely to belong to the same objects, whereas disconnected parts usually belong to different objects (Roelfsema & Singer, 1998). In the laboratory, contour tracing is studied by a task where observers are required to determine whether two dots lie on the same contour in a pattern consisting of two (or more) intermingled contours. A typical finding is that the time it takes to provide an answer increases monotonically, but not linearly, with the distance between the dots on the contour.

In their pioneering study, Jolicoeur et al. (1991) devised simple stimuli consisting of a set of parallel straight or curved lines to isolate relevant factors that contribute to dynamics of tracing: the distance between the target and distractor contours and the amount of curvature in the curved contours. Their study revealed that: (a) the tracing time increases monotonically and roughly linearly with the length of the contour, (b) the tracing speed decreases with decreased spacing between the target and distractor contours, and (c) the tracing speed decreases with increased contour curvature. These findings help to explain why the tracing time was not a linear function of the distance in studies employing two contours that wiggle around each other. In such stimuli, the distance between the target and distractor contours, as well as their curvature, vary considerably along the path that needs to be traced. Therefore, in this part of the thesis, the modeling efforts will be on Jolicoeur et al.'s (1991) results.

To simulate properties of mental contour tracing, the F-WTA network was enriched with the extended lateral connectivity. Hence, the neural model of contour tracing consists of three components: The F-WTA network, the contour detection network (CDN), and the L-junction detection network (LDN). It was assumed that there were separate F-WTA networks for each orientation, so that activity spreading in one orientation would not jump to other orientations at contour junction. For the computational simplicity, only horizontal and vertical orientations were employed. Furthermore, it was proposed that F-WTA nodes do not mutually interact directly but were engaged in a feedback loop with the same-oriented multi-scale CDN. In other words, lateral excitatory interactions in the F-WTA network were multiplicatively gated by the output of the CDN.

The proposed model of the CDN network consists of two orientations: horizontal and vertical, and four spatial scales: tiny, small, medium, and large, denoted as 1, 2, 3, and 4, respectively. The input image is convolved with a set of multi-scale, even-symmetric Gabor filters to simulate oriented receptive fields known to exist in the visual cortex (Daugman, 1980; Marčelja, 1980). The output of the convolution is denoted as *c*-units. The model of *a*-units receives scale-dependent feedback projections from the F-WTA network with the same orientation preference (e.g., horizontal, or vertical). The lateral extent of the feedback projections to the *a*-units scales with the size of the receptive field of the corresponding *c*-units. Furthermore, the output of *a*-units is multiplicatively gated by *c*-units before being returned to the F-WTA network. The distinction between *a*-units that receive attentional signals from the F-WTA network, and *c*-units that are unaffected by attention is reminiscent of the distinction between the A- and N-units found in the visual cortex (Pooresmaeili et al., 2010). Importantly, the role of CDN is to open or close feedback loops between neighboring nodes in the F-WTA network in a scale-dependent manner, but the F-WTA network is the carrier of the activity spreading related to attention.

Since the interaction between F-WTA and CDN explains dynamics of contour tracing within a single orientation, the LDN model is designed to explain how contour tracing is engaged in cross-orientation interactions at L-, T-, or X-junctions. LDN allows the activity spreading across orientations at L-junctions, but not in other types of junctions. Empirical support for the existence of such corner detectors is provided by Anzai et al. (2007). LDN is composed of two *e*-units that exhibit enhanced response at contour ends and corners, and one *l*-unit that detects superposition of perpendicular *e*-units. It is assumed that both types of these units exist at the smallest spatial scale only. Furthermore, *e*-units receive input from the same-oriented *c*-units. The *l*-unit has two dendrites, each receiving input from a distinct *e*-unit at the same location. The threshold of the *l*-unit is set in a way so that it operates as an AND gate, i.e., it is activated only when it receives suprathreshold signals from both *e*-units. Hence, the *l*-unit will selectively respond to the L-junction, but not the T- or X-junction. The output of the *l*-unit establishes a link between the horizontal and vertical F-WTA networks via the same multiplicative gating as that used in the CDN. Importantly, the LDN network can be generalized to a model with many orientations in such a way that the *l*-unit receives input from each *e*-unit on a separate dendrite.

A set of computer simulations was performed to illustrate that the proposed model could account for the effect of distractor proximity, the effect of contour curvature, and the effect of object-based attentional cueing. Moreover, simulations were performed to show how the F-WTA network handles gaps at the contour or contour intersections, solves the spiral problem, and separates the representation of occluding from occluded object. The robustness of the model was also verified by examining the extent to which its behavior is affected by changes in network parameters.

2.1.2.2. The Effect of Distractor Proximity

Jolicoeur et al. (1991) found that the tracing speed is modulated by the proximity of the target and distractor contours, that is, the tracing becomes increasingly slower as the proximity is increased. To simulate these findings, we employed an input image consisting of vertical lines that were horizontally separated by a variable number of pixels. Results showed that the output of the vertical *c*-units correlated with the contour spacing. Contours were encoded only at the tiny scale (c_1-units) when the horizontal separation between the contours was only one pixel wide. Scale-dependent flanking inhibition of Gabor filters reduced the activation of the c_2-, c_3-, and c_4-units, and scale-dependent thresholds T_s made these weak activations subthreshold. The exception to this inhibitory effect were the c-units positioned at the very edge of the grating. They survived thresholding even at lager scales because they received weaker total inhibition. Next, when the horizontal separation between the contours was two pixels wide, the c_2-units were released from inhibition along with c_1-units. Similarly, when the horizontal separation between the contours was three pixels wide, the c_3-units were released from inhibition alongside with c_1-, and c_2-units. Finally, when the horizontal separation between the contours reached the maximum of four-pixels width, the c-units enabled the representation of contour at all four spatial scales.

The effect of distractor proximity on dynamics of the vertical F-WTA was demonstrated by showing snapshots of evolving F-WTA activity taken at five representative time points in response to the four input configurations described in the previous paragraph. The five representative time points were: presentation of the spatial cue (t = 300 ms), removal of the cue and the beginning of the activity spreading (t = 500 ms), enhancement in the activity spreading along the target contour (t = 600 ms, and t = 800 ms), and the end of the simulation (t = 1,500 ms). At the beginning of the simulation all contours were selected because of the same input amplitude. It was not possible to automatically generate tracing within the network. Thus, tracing started after applying an external spatial cue in the middle of the image (Crundall et al., 2008). That is, the attention-related activity spreading is not an obligatory process and its engagement depends on the task demands (Drummond & Shomstein, 2010; Shomstein & Yantis, 2002, 2004). After removal of the spatial cue, the network did not return to the initial state where all contours were selected. Instead, neural activity started to propagate from the cued location to the end of the contour without spillover to the background. This was a consequence of the lateral excitatory interactions mediated by CDN, which was described earlier.

It is important to notice that the F-WTA network automatically adjusted the speed of tracing depending on the proximity of distractor contours. The activity spreading was the slowest (tracing ended at about t = 1,500 ms) when the distance between the contours was one pixel wide, and the fastest (tracing ended at about t = 600 ms) when the distance between the contours was three and four pixels wide. In this last condition, the F-WTA network received input from a-units of all scales because of the suprathreshold activity in c-units at all scales. Such multi-scale activation increased the size of the integration zone of F-WTA nodes and consequently considerably increased the speed of tracing. Here, Marić and Domijan (2019) emphasized that the model implements object-based attention by allowing the neural activity to spread along the whole object, rather than to move along the contour (Houtkamp et al., 2003; Roelfsema et al., 2010; Scholte et al., 2001). This was achieved by keeping the old nodes, which were already activated in the earlier stages of tracing, active with the new nodes appended to the existing representation of contour until the end of tracing. In the background of these findings is the retrograde signaling mechanism that prevents lateral inhibition among winning nodes, irrespective of their total number.

The same simulation of the effect of distractor proximity was also examined from the perspective of the vertical F-WTA nodes lying on the target contour with a target-distractor distance from one to four pixels. Nodes positioned on the target contour were 4, 8, 12, and 16 pixels away from the cued node. Results illustrated the same dynamics as described in the previous paragraph. The cued location was the only winner when the cue was on, and tracing proceeded by the sequential excitation of the connected nodes after the removal of the cue.

Finally, a comparison between the results of simulation and behavioral data (Experiment 3 of Jolicoeur et al., 1991) was made. Two empirically observed trends were detected in the model's output. First, the tracing time increased monotonically with the distance from the cue to the target location. Second, the slope of the tracing time curve increased as a function of the increased proximity between the target and distractors.

2.1.2.3. The Effect of Contour Curvature

The second effect that was simulated was the effect of contour curvature. When the distractor proximity was fixed, Jolicoeur et al. (1991) found that tracing becomes increasingly slower as the curvature on the contour increases. To simulate these findings, a low-resolution input image was employed consisting of a small horizontal displacements of vertical contour segments to emulate the contour curvature. That is, the input was the vertical grating with large, medium, and small contour curvatures, or straight contours, denoted as A, B, C, or D, respectively. Stimuli with large contour curvatures contained contour segments that were displaced horizontally three times to the right and then two times to the left to represent a curved line. Similarly, medium curvature contours had two displacements to the right and one to the left, and small curvature contours had only one displacement to the right and one to the left. Displacements were always made by one pixel. The spacing between the contours was kept constant. Results showed that the output of vertical *c*-units correlated with the contour curvature. The stimulus with large contour curvatures produced short contour segments that activated only c_1-units. In the second condition, the stimulus with medium contour curvatures had longer contour segments that activated c_2-units along with c_1-units. In the third condition, contour segments were large enough to additionally activate c_3-units along with smaller spatial scales. Finally, when the input was composed of straight contours, all four spatial scales of the CDN were activated in parallel.

Simulation of the effect of contour curvature on dynamics of the vertical F-WTA was illustrated by snapshots of evolving F-WTA activity taken at five representative time points in response to the four input configurations depicted previously. As in the previous simulation of distractor proximity, the spatial cue was applied to the top of the contour placed in the middle of the input image. The cue enabled the network to behave similarly to a standard WTA network and to select only the cued location. After the removal of the spatial cue, the network started to trace contour by selectively amplifying the activity of neighboring units (at about t = 500 ms across all rows). Importantly, results showed that the speed of tracing was modulated by the response of c-units. The activity spreading was the slowest (about t = 1,300 ms) in the large contour curvature condition, and the fastest (about t = 800 ms) in the straight contour curvature condition.

The effect of the contour curvature was further examined by showing the temporal dynamics of the vertical F-WTA nodes lying on the target contour with a degree of contour curvature ranging from large curvatures to straight lines. Nodes positioned on the target contour were 8, 16, and 24 pixels away from the cued node. At the beginning of the simulation, the cued location was the only winner when the cue was on. After the removal of the cue, contour tracing proceeded by the sequential excitation of connected nodes. Tracing ended when it reached the last pixel of the target contour. This occurred at different time points depending on the contour curvature.

Finally, a comparison between the model's output and behavioral data (Experiment 1 of Jolicoeur et al., 1991) was made. Again, two empirically observed trends were detected. First, the tracing time increased monotonically with the distance from the cue to the target location. Second, the slope of the tracing time curve increased as a function of the increased contour curvature. That is, contour tracing was increasingly slower as the amount of curvature in the curved contours increased. Interestingly, the speed of tracing was generally slower when compared to the previous simulation with the same parameter set. This was a consequence of imperfect alignment of curved contour segments with the excitatory part of Gabor filters, so c-units were not maximally activated as in the simulation of the effect of distractor proximity.

2.1.2.4. The Effect of Object-Based Attentional Cueing

Object-based cueing involves temporary brightening (cueing) one of the contours in the grating. In such a task, the cue shortly appears on the target contour following by the presentation of dots whose connectedness should be established. Here, it is important to emphasize that the cue disappears before the dots appear. Therefore, participants need to store the locations of the target contour in short-term memory to effectively use the cue when they attempt to solve this task. McCormick and Jolicoeur (1992) investigated the relationship between attentional cueing and mental contour tracing and found that an object-based cue substantially reduced the speed of tracing. Results suggested that all segments of the cued contour were selected together, so no tracing was required.

In the simulation of the effect of object-based attentional cueing, we investigated the activation of the vertical F-WTA network in response to a temporary increase in the input amplitude of the target contour. In the first simulation, when the cue duration was 200 ms, F-WTA successfully selected the cued contour and stored it in working memory. In the second simulation, when the cue was presented for much shorter interval (50 ms), as in the study of McCormick and Jolicoeur (1992), F-WTA failed to segregate the target contour from distractors and returned to initial state in which all contours were selected together. In this simulation the input amplitude was kept the same as in the previous simulation. Finally, in the third simulation when the cue amplitude was increased, the network regained its ability to segregate the target contour from distractors over the cue duration of 50 ms.

2.1.2.5. Other Tracing Effects

Gaps on the contour may disrupt tracing (Ullman, 1984), so it is important to establish to what degree the proposed model is immune to such disruption. To address this issue, we ran a simulation with a vertical grating as the input image (stimuli like those used in the simulation of the effect of distractor proximity) to illustrate how the F-WTA network handles one-pixel and two-pixel gap on the contour. The one-pixel gap was placed near the starting point of tracing, and the two-pixel gap was placed farther away from the cued contour. Results showed that when contours were closely spaced, the activity enhancement in F-WTA could not cross the one-pixel gap. In the condition with larger spacing among the contours, the activity enhancement successfully crossed the one-pixel gap, but got stuck at the second, larger gap. Finally, when the spacing between the contours was increased even further, the activity enhancement in F-WTA crossed both gaps and completed the spatial representation of the target contour. Importantly, F-WTA may cross even wider gaps, but this would require the recruitment of even larger spatial scales (e.g., c_4-units or lager).

Up to now, we considered only simple stimuli such as gratings. However, to be useful in everyday situations, tracing should be applied to more complex patterns composed of vertical and horizontal contour segments. Here, we employed the famous spiral problem (Minsky & Papert, 1988) to demonstrate how the F-WTA network solves it. The input pattern was composed of one or two spirals. The outputs of vertical and horizontal *c*-units were combined in a single image by computing their maximum at each pixel. This maximum was displayed in the image. Like in the previous simulations, the external spatial cue was applied and then withdrawn to initiate tracing. In the simulation where the input pattern was composed of one spiral, the activity enhancement progressively spread along the cued contour from t = 500 ms until the end of the contour shortly before t = 4,000 ms. The steady state activity of F-WTA suggested the presence of a single connected spiral. On the other hand, in the simulation where the input pattern was composed of two spirals, the activity enhancement progressed in a similar way, but it reached the end of the cued contour before t = 2,000 ms. The steady state activity of F-WTA suggested the presence of two separated spirals. Importantly, the activity spreading successfully switched between perpendicular orientations at corners of spirals due to the activation of *l*-units.

The next question that we addressed is how the network will trace patterns with X- or T-junctions, i.e., contour intersections. Without orientation-specific mechanisms that constrain tracing, the tracing would spill over in all directions. Human observers tend to choose the direction that requires the smallest change in orientation, which is closely related to the well-known Gestalt principle of good continuation (Brooks, 2014). To this end, we designed *l*-units in such a way to selectively respond to the L-junction but not to the X-junction.

The simulation of the contour tracing across X-junctions demonstrated that the F-WTA network spread activity enhancement according to the Gestalt principle of good continuation. The input consisted of two intersecting squares, and the cue was applied on the upper left square. After the removal of the cue, the activity enhancement in the horizontal F-WTA network began to spread simultaneously to the left and to the right from the cued location. The vertical F-WTA network could not exhibit activity spreading at the cued location as it did not receive the input from the vertical *c*-units. However, the enhanced activity in the horizontal F-WTA network crossed L-junctions at the corners of the square and activated the vertical F-WTA network, which continued to propagate enhancement in the vertical direction. Next, the activity enhancement passed through X-junctions and, importantly, there was no activity spillover to the horizontal F-WTA because *l*-unit was inactive here. At the end of the simulation, both networks reached a steady state with the active representation of the cued square.

Next, the simulation of the contour tracing across T-junctions showed that the F-WTA network correctly separated the representation of the occluding from the occluded object. Which object is selected depends on the location where the cue was applied. Here, the same effect was observed as in the simulation described in the previous paragraph. When the occluding object was cued, tracing proceeded along its contour without spillover to the contour of the occluded object at the T-junction, and vice versa when the occluded object was cued. Interestingly, even though the F-WTA network did not have the capacity to represent different depth planes for occluding and occluded objects, it properly detached them during tracing.

2.1.2.6. Discussion

The model of object-based spatial selection developed in this study shares similarities with the zoom lens model proposed by McCormick and Jolicoeur (1991, 1994), which was developed to account for behavioral findings on contour tracing. According to the zoom lens model, the spotlight of attention narrows or widens its radius depending on the spatial resolution that one wants to achieve. McCormick and Jolicoeur (1991, 1994) have identified five component processes that a contour tracing operator should have: 1) contour detection process, 2) zooming process, 3) process of determining whether the second contour segment has been located, 4) computation of the direction of the next attentional shift, and 5) shifting to a new region process. Since the zoom lens model is purely descriptive, the present model of object-based spatial selection provides a specific implementation of neurocomputational mechanisms underlying described attentional processes. The *c*-units implement the first process, while the multiplicative gating of *c*- and *a*-units implements the second process. Last three processes are intrinsic to the F-WTA network, that is, F-WTA automatically finds a path to spread activity along the contour without spillover to the empty space around it. In addition, LDN regulates activity spreading along L-, X-, and T-junctions, and prevents spillover to distractor contours.

The F-WTA network shares features with the filling-in model of brightness perception (Grossberg & Todorović, 1988). In this model, a neural substrate of surface perception is the diffusion of electrical activity among neighbors, which allows brightness or color signals to fill in the interior of a surface because the diffusion is blocked at surface borders. However, diffusion spreads slowly without the ability to adjust the speed of activity spreading. It has no ability to sustain activity in short-term memory related to the target object after the input vanishes. In contrast, the F-WTA network solves these problems by incorporating the multiplicative gating of excitatory interactions and self-recurrent collaterals. In addition, the F-WTA is sensitive to new inputs and even to abrupt visual onset of new objects, as it was shown in Marić and Domijan (2018a). Saturation of the node's activity is prevented by dendrites, so F-WTA makes smooth transition from old to new inputs.

At the end, there are some limitations of the proposed F-WTA network. One problem is that its pattern separation capability is too strong because the distractor contour was suppressed to zero, while the target contour remains active as it was presented alone. However, neurophysiological data provided by Roelfsema et al. (2003, 1998) suggest that the activity difference between neurons encoding target and distractor contours is relatively small. However, they tested only neurons in V1. Bogler et al. (2011) found that PPC is involved in the WTA computation in a way that it combines feature selectivity with spatial attention to compute a feature specific priority map (Veale et al., 2017). Consequently, PPC may contain the representation of the target contour similar to the output of F-WTA. There is also a possibility that F-WTA may be located in the pulvinar nucleus of the thalamus (Zhou et al., 2016). Irrespective of its exact anatomical location, we proposed that the F-WTA network is a part of the attentional network dedicated to target selection and distractor filtering that may operate independent of the network dedicated to visual perception.

In summary, this study demonstrates that when the F-WTA network is embedded in a larger multi-scale architecture for contour and L-junction detection it is capable of attentional labeling of connected image parts. Dynamics of this labeling is consistent with the empirically observed dynamics of mental contour tracing. The proposed model offers a neural interface for the interaction between visual perception and cognition, and for the implementation of incremental grouping of image elements. These findings suggest that the second part of the first problem of the thesis is successfully addressed.

2.1.3. Literature Review on Cognitive Penetrability of Vision

The nature of the relationship between visual perception and cognition remains an open and important question not only in the field of psychology but in neuroscience, philosophy of mind, artificial intelligence, psychiatry, and aesthetics. The assumption that visual perception is cognitively penetrable implies that cognitive processes such as thinking, reasoning, expectations, beliefs, emotions, values, etc., can directly influence and change the content of visual perception. In terms of information processing theory, this means that top-down processes can directly influence bottom-up pick up of a sensory information (Palmer, 1999). However, if the visual system is cognitively penetrable, cognition may interfere with the main task of vision. That is, to create accurate mental representation of the observers' environment. To what extent vision is functionally independent of cognition and emotions and how is our behavior coordinated with physical features of the external environment are theoretically and empirically important questions that require new theoretical approaches to resolve them. Hence, the aim of the Marić and Domijan (2018b) review paper was to offer an overview of theoretical arguments for and against the cognitive penetrability hypothesis, as well as empirical findings that support or refute it.

Neuroscientific data have shown that vision is composed of a set of parallel and hierarchically organized neural networks specialized to analyze different features of visual stimuli such as color, motion, orientation, etc. (Lennie, 1998). Therefore, when analyzing the problem of cognitive penetrability in vision, it is necessary to take into account neuroscientific findings about the interactions between cortical areas involved in vision. Visual signals travel from the retina to the LGN, where the segregation of visual pathways into magnocellular and parvocellular streams is further continued in the first visual receiving area – the primary visual or striate cortex (V1) (Callaway, 2005). Projections from the striate cortex to extrastriate visual areas (V2, V3, V4, and MT) can be further divided into two functionally segregated visual pathways or streams – a ventral stream and a dorsal stream. The ventral stream processes information about object identity, whereas the dorsal stream processes information about object location and contributes to visually guided movements (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). The ventral stream projects toward the ventral temporal cortex, and the dorsal stream projects toward the parietal lobe. The functional separation between streams is not strict and there is evidence for the interaction between them (Bell et al., 2014).

The standard model of the neural basis of vision relies on the assumption that the visual system is a hierarchical structure. Visual processing starts with the representation of elementary features (e.g., color, motion, orientation, …) at lower levels of the hierarchy (V1) and continues to their generalization and integration into more complex shapes, categories, and objects at higher levels of the hierarchy, such as the inferotemporal cortex (IT). However, visual processing is much more complex than a hierarchical or feedforward view suggests since there is a dense network of interconnected feedforward, lateral, and feedback pathways (Felleman & Van Essen, 1991). Feedback projections have been found at all stages in the visual cortical hierarchy and are more numerous than the feedforward connections (Ahissar & Hochstein, 2004; Gilbert & Li, 2013; Hochstein & Ahissar, 2002). These observations suggest that visual processing is bidirectional and subject to top-down influences rather than strictly hierarchical.

Feedback connections have different anatomical and physiological properties than feedforward connections (Markov et al., 2013; Bastos et al., 2015; Michalareas et al., 2016). Macknik and Martinez-Conde (2009) suggested that feedback modulates neural activity in visual perception rather than drives it, with its clearly established important role in attentional mechanisms. They argued that the main role of feedback connections is to facilitate and suppress feedforward signals as a function of attentional load.

Pylyshyn (1999) introduced the concept of cognitive impenetrability of visual perception starting from the Fodor's (1983) modularity of mind hypothesis. According to the Fodorian view of mental architecture, the human mind is organized much like computer architecture with central processor and a set of specialized modules responsible for executing specific tasks. The modules are, for example, vision, audition, motor control, etc. Modules operate independently of each other and of central processor in order to efficiently perform their dedicated role. They cannot influence each other. The central processor only receives the output from the modules but cannot affect their internal operations. Consequently, computations that take place inside the module are concealed or informationally encapsulated from cognition. Pylyshyn (1999) has argued that visual perception is Fodorian module with fixed, innate architecture. Its operation is based on a domain-specific set of principles that are fundamentally different from principles governing the operations of the central processor that implements general cognitive functions, such as thinking and reasoning.

A part of visual processing is cognitively penetrable, and it is related to object recognition. Thus, Pylyshyn (1999) made a distinction between visual processes that are cognitively penetrable and other set of processes that are not cognitively penetrable. Following Marr (1982), Pylyshyn (1999) referred to a part of visual perception that is encapsulated from cognition as an *early vision*, and visual perception that is susceptible to cognitive influences as a *late vision*. The early vision is defined functionally, without reference to its anatomical locus that is not known exactly. It encompasses processes from light registration in photoreceptors through analysis of perceptual features to the computation of tridimensional (3-D) representation of the surfaces of objects in the environment. Semantic information starts to influence visual processing only after the early vision determined 3-D structure of the environment by building representation adequate for object recognition, which is a part of the late vision.

2.1.3.1. Arguments for Cognitive Impenetrability of Vision

Apart from theoretical arguments reviewed above, Pylyshyn (1999) has offered several empirically grounded arguments for cognitive impenetrability of visual perception. First, knowledge about the existence of perceptual illusions (such as the Müller-Lyer illusion) does not make them disappear. That is, previous knowledge that our visual perception does not exactly match the physical situation in the environment still does not alter the content of our visual experience.

Second, there are many regularities in visual perception that depend only on the properties of the visual input and are automatically extracted from it. Principles of visual perception differ from the principles of thinking and reasoning, that is, of the central processor. For example, Gilchrist et al. (1999) have identified the rule known as a highest-luminance-as-white by which we assign white color to the surface that emits the largest amount of light in a given stimulus. This rule cannot be applied and does not make sense in domains outside of visual perception. Also, it is not a general rule according to which the central processor can operate.

Third, neuropsychological studies have observed the partial independence of vision and other cortical functions. For example, people who suffer from visual agnosia, (i.e., an impairment in object recognition), do not show simultaneous deficits in general cognitive functioning (Farah, 1990). On the other side, people who suffer from deficits in reasoning and thinking, in principle, do not show simultaneous impairments in visual perception.

Fourth, empirical evidence supporting penetrability of vision can be reinterpreted in a way that top-down modulations affect a post-perceptual stage of processing or they affect directing spatial attention that operates before the early vision. For example, knowledge about the location of interesting object in the visual scene can influence how we will direct spatial attention before the early vision starts to encode visual scene. In addition, the knowledge can also affect decision-making processes that take place after the stimulus encoding in the early vision is finished.

Pylyshyn’s arguments against penetrability of vision were further discussed and elaborated by Raftopoulos (2009, 2014). Raftopoulos (2009, 2014) specifically focused on electrophysiological findings about the speed on the signal flow rate through the visual system to clearly determine the temporal boundary between early and late vision (Lamme, 2003;

Lamme and Roelfsema, 2000; Roelfsema, 2005). According to these findings, visual representation of stimuli is created in three processing steps. The first step takes place within the first 100 ms after stimulus onset when signal transmission progresses from the retina through the visual cortical areas to IT cortex. This signal transmission is unconscious and resistant to feedback from higher-level areas in visual hierarchy. Therefore, it is considered a pure bottom-up form of information transmission in the visual cortex. In the second step that takes place about 120 ms after the stimulus presentation, the early visual areas and higher extrastriate areas engage in local recurrent processing, which are necessary for conscious visual experience. Raftopoulos (2009, 2014) argues that the first two steps are the processes that correspond to Marr's concept of the early vision and are not directly influenced by cognition. That is, signal transmission during early vision is not affected by top-down signals produced in cognitive areas but is restricted within the visual areas of the brain. Raftopoulos (2009) emphasized that only the third step of visual processing, which takes place between 150 and 200 ms after the stimulus presentation, is cognitively penetrable. In this step signals from the frontal and parietal areas and mnemonic circuits in the hippocampus begin to modulate perceptual processing in the visual cortex.

2.1.3.2. Arguments for Cognitive Penetrability of Vision

Concepts of cognitive impenetrability of vision and modularity of mind runs contrary to a *New Look movement* in perception and experimental findings that supported it (Bruner, 1957). According to this view, values and needs affect all levels of perception and determine how we perceive the environment. Perception is inferential process, so it can be thought of as a form of problem solving. The input that is registered on the retina is impoverished and ambiguous. In a sense, raw sensory information represents a problem to be solved. The problem of finding the best interpretation of what is given in the input image is solved by supplying top-down information in a form of a hypotheses or predictions about the likely causes of registered sensory information. From these considerations it follows that there is no real distinction between perception and thought (Bruner, 1957).

Recently, New Look movement regained its popularity in part because of the development of new theoretical constructs such as predictive coding and Bayesian models of perception (Clark, 2013; Friston, 2010; Hohwy, 2013; Lupyan, 2015a; O'Callaghan et al., 2017; Vetter & Newen, 2014). Predictive coding is a biologically and computationally plausible description of information processing in the brain. The main hypothesis of predictive coding is that, during visual perception, the brain generates top-down predictions that are derived from internal model of the environment. Predictions facilitate, guide, and constrain the processing of incoming sensory input (Friston, 2010; Rao & Ballard, 1999). The internal or generative model of the environment is built from past experiences stored in memory, contextual associations, as well as sensory information arriving from other sensory modalities. Here, it is important to emphasize that we observe objects in typical conditions in which they appear together with other objects, and their cooccurrences allows us to learn the associations between them. Contextual associations then become a rich source of information that aids the process of object recognition (Bar, 2004). An important implication of predictive coding is that vision is an interactive, two-way process of merging bottom-up and top-down signals. In other words, the brain never perceives the environment as a raw sensory image. Instead, it always starts from some assumption about what environment might look like. Therefore, the predictive coding model is a new theoretical formulation of cognitive penetrability of visual perception.

Marić and Domijan (2018b) also reviewed neuroscientific evidence in favor of cognitive penetration in vision. For example, Vandenbroucke et al. (2016) employed functional magnetic resonance imaging (fMRI) to show that object knowledge can create expectations that alter activity in visual cortical areas representing elementary perceptual features, such as color and shape. During perception of achromatic objects with diagnostic colors, for example a banana or tomato, they managed to decode brain signals evoked by color-selective cortical areas. Object knowledge modified early visual areas, that is, penetrated the cortical area V4 involved in color perception. Results suggested that observers indeed perceive a typical color (yellow) while observing gray banana. In addition, there is also neuroimaging evidence that social and emotional knowledge can also modulate visual processing.

A detailed overview of a number of behavioral studies suggesting cognitive penetration in vision was provided by Firestone and Scholl (2016). Marić and Domijan (2018) focused their review on studies that caused the most controversy and motivated the discussion on methodological problems arising in studies of cognitive penetrability of vision. For example, Bhalla and Proffitt (1999) examined the relationship between conscious slant perception and individual’s physiological potential. In four experiments they showed that hills appeared steeper to people who were encumbered by wearing a heavy backpack, were fatigued, were of low physical fitness, or were elderly and/or in declining health. Similarly, Witt et al. (2004) showed that the perceived distance was a function of actual distance as specified by optical variables, perceiver's intention to do the action, and the effort associated with this intended action. They concluded that perception of distance combines the perceived geometry of the environment with individual's behavior goals and the potential of body to achieve these goals. That is, if perceiver intents to walk to a target, then he will see distance in terms of the effort required to walk to it. Furthermore, the consequence of the glucose consumption after starvation is that the hill slant was perceived as less steep (Schnall et al., 2010), and distances were perceived as shorter (Cole et al., 2013) than after non-caloric sweetener consumption.

Cognitive and emotional effects on visual perception are not limited to perception of spatial features only but extend to brightness and color perception. Banerjee et al. (2012) explored whether recalling positive or negative concepts such as evil (as exemplified by unethical deeds) and goodness (as exemplified by ethical deeds) would influence the perception of brightness of the room. Results showed that participants in the unethical condition judged the lab room to be darker than did participants in the ethical condition. Moreover, participants who recalled their own past unethical deeds preferred products that would make the room brighter and thereby to reduce the negative feelings associated with darkness. Along the same line, Song et al. (2012) found that people judged smiling faces as perceptually brighter than frowning faces. In four studies they demonstrated that this effect appeared regardless of the type of assessment (a binary choice task or an absolute judgment task) and the type of the stimuli (schematic faces or real faces).

2.1.3.3. Methodological Aspects of Cognitive Penetrability of Vision Research

Results that inspired the New Look movement were not replicated in later studies that identified and controlled for a number of intervening variables. This suggests insufficient experimental control and lack of methodological rigor in early attempts to establish the evidence for cognitive penetrability of vision. Firestone and Scholl (2016) suggest that similar methodological problems arise in modern research too. They provided a detailed analysis of the methodology of studies that support the penetrability of vision. They concluded that the effects observed in these studies could be explained without invoking the influence of cognition or emotions on vision. The most important contribution of their work was highlighting the need to improve scientific methodology and control over relevant variables in the design of the study. In particular, they identified six methodological pitfalls that prevent us from reaching any firm conclusion regarding cognitive penetrability of vision. Future studies will need to address these issues before they may claim that they found evidence for cognitive penetration of vision. Short description of each pitfall is provided below.

The first pitfall is an overly confirmatory research strategy. In general, the experimental hypothesis can be tested in two ways: by finding an effect that is consistent with the theory, and by the absence of an effect when the theory predicts its absence. Cognitive penetrability studies focused exclusively on the first approach. Therefore, Firestone and Scholl (2016) suggested that many potentially important findings are missing from the literature, and they may help in resolving penetrability debate. A specific example of the distinction between confirmatory and disconfirmatory hypotheses is seen in the research by Firestone and Scholl (2014) who were inspired by an interesting anecdote in art history known as *the El Greco fallacy*. Famous painter El Greco is known for his tendency to paint the unusually elongated human figures. Based on this tendency, art historians hypothesized that he suffered from astigmatism that stretched his perceived environment along the vertical axis. However, if he really experienced the world in an elongated way, he would also choose elongated canvases so that the traces of the assumed distortions would not be visible in the paintings themselves. However, this was not the case as he used similar canvases as other artists of his time. Thus, it is more likely that the elongated figures in his paintings are an element of his painting style, and not distortions of his perception (Anstis, 2002; Firestone, 2013).

A similar error may occur when interpreting findings on top-down effects in vision. For example, Firestone and Scholl (2014) showed that the effect of action capability on spatial perception (Stefanucci & Guess, 2009) and the effect of emotional valence on lightness perception (Banerjee et al., 2012) appeared even in a condition where they should not. For example, when participants made lightness judgments on a lightness scale (by picking a shade of gray from a range of grays) instead of on a numerical-report scale they still produced the difference in lightness judgments between positive and negative emotions. This was the case despite the fact that any perceptual distortion should cancel out because lightness scale is present in the same room whose lightness is being judged. That is, any darkening (or lightening) of the room that purportedly arise from emotional states (positive vs. negative) should also darken (or lighten) the scale on which judgments are made, so no lightness difference should be observed. This suggests that the reported effects are not perceptual but arise at some post-perceptual stage of processing (e.g., judgment or decision making). Therefore, Firestone and Scholl (2016) recommended that both confirmatory and disconfirmatory research strategies need to be applied concurrently if one wish to present a strong case for the cognitive penetrability of vision.

The second pitfall is a failure to distinguish perception from judgement. Because it is difficult to draw the line between judgment and visual perception, it is possible that the results supporting cognitive penetrability are due to top-down effects on judgment rather than on vision. Therefore, future research will have to consider the difference between vision and judgment and empirically separate them to avoid misinterpretations of the findings (Firestone & Scholl, 2016). The results of Witt et al. (2004) on the effect of physical effort on the distance perception are the example of this pitfall.

The third pitfall refers to demand characteristics of the experimental task and response bias. Vision experiments are conducted in controlled laboratory settings that allow for a maximal control of confounding variables. However, experiment also involves social interaction between the participant and the experimenter. Participants are not passive recipients of task instructions. They actively try to understand the true purpose of the experiment by observing demands imposed by the experimental task. They also try to behave as *good* participants and to please experimenter by adjusting their responses along the hypothesis they assume is being tested. Therefore, task demands pose an important threat to the internal validity of experiment (Shaughnessy et al., 2012). To address this threat, researchers need to invest a great deal of effort to hide all signs that indicate the aim of the study. Klein et al. (2012) warned that many contemporary studies do not pay enough attention to the problem of demand characteristics leading to a situation where it is difficult to discover the real effects. This also contributes to the problem of replicability in psychological research.

An example of a study where demand characteristics played the important role in contaminating the research findings is the study of Bhalla and Proffitt (1999) who showed that wearing heavy backpacks led participants to overestimate the steepness of the hill. This finding suggests that kinesthetic information provided by the muscles modulates perception of slant in accordance with the cognitive penetrability hypothesis. However, the effect of backpack may simply reflect experimental demands because it was obvious to participants what is the hypothesis that has been tested. When the relationship between backpack and perceptual task is made less obvious, by inventing a cover story that directed participants' attention to other aspects the task, the effect of backpack on slat perception disappeared (Durgin et al., 2009).

The fourth pitfall are low-level differences in stimulus that refer to manipulations of the stimuli used across experimental conditions. For example, the intended top-down manipulation can be confounded with the changes in low-level visual features of stimuli so that they may drive the perceptual differences across conditions. Firestone and Scholl (2016) suggested that it is important to separate high-level from low-level variables in a way to preserve the high-level factor, while eliminating the low-level factor, or vice versa. For example, Firestone and Scholl (2016) replicated Levin and Banaji (2006) study who found that Black faces look darker than White faces, even when matched for mean luminance. Firestone and Scholl (2016) used the blurred versions of the face stimuli to eliminate race information while preserving low-level differences in the images. After blurring, the vast majority of observers claimed that the two faces actually had the same race, but they nevertheless judged the blurry image derived from the Black face to be darker than the blurry image derived from the White face. This finding implicates the action of some low-level variable rather than a high-level variable such as race.

The fifth pitfall is related to peripheral attentional effects. Changing what we see by selectively attending to different locations is similar to changing what we see by moving our eyes. In both cases, we choose the input to vision. However, in both cases, the influence of attention is completely independent of the reason for directing attention, i.e., attention is not sensitive to the content of that intention or belief. Thus, such influence cannot be considered as the cognitive penetration of vision (Firestone & Scholl, 2016). On the other hand, Lupyan (2017a) argues that some other forms of attention, such as feature-based attention or attention to semantic categories, do indeed change the content of visual perception in a way that change the appearance of the object. In doing so, he offered several demonstrations in which additional information about what is given in the image dramatically changes the perception of that image. Therefore, it is important in future research to control or measure the effect of directing spatial attention to distinguish it from other top-down effects. Also, it will be important to distinguish whether feature-based attention or attention to semantic categories has different impact on visual perception when compared to spatial attention.

Finally, the sixth pitfall is failing to take into account the effect of memory on recognition. In many studies the effect of cognitive penetration is confused with the recognition of stimuli. For example, studies showed that assigning linguistic labels to simple shapes speeds up the reaction time in visual search and other recognition tasks (Lupyan & Spivey, 2008; Lupyan et al., 2010; as cited in Firestone & Scholl, 2016). However, in addition to visual processing, recognition also involves memory necessary to compare a given visual stimulus with the remembered representation. Since visual recognition incorporates both perception and memory, the recommendation for future research is to distinguish between these two processes (because memory effects have no implications for the nature of perception) if one wants to demonstrate the effect of cognitive penetration on vision (Firestone and Scholl, 2016).

2.1.3.4. Discussion

In summary, it can be concluded that no consensus has been reached on whether there is an impact of cognition and emotions on visual perception. Further theoretical work is needed to clarify the role of feedback connections in the visual cortex in generating top-down effects observed in behavioral studies. In contrast to predictive coding and Bayesian models, Grossberg's (2013) adaptive resonance theory offers an alternative view on the feedback connections and their role in the cortical information processing. On this view, feedback connections do not alter bottom-up input. Instead, they help to stabilize learning and memory because they read out learned expectations that are compared with the bottom up input. However, it is not clear how this theory would explain the findings that support the cognitive penetrability of vision. Therefore, further theoretical work is needed to explain at what stage of processing within the adaptive resonance theory the interaction between perception and cognition occurs.

Future empirical studies will need to take into account the recommendations put forward by Firestone and Scholl (2016) to distinguish the actual effects of cognition and emotions on vision from the effects of confounding factors present in the experiment. Even if stringent control of confounding variables is achieved, the question remains as to what extent it is possible to separate visual perception from subjective interpretations and judgments that are often conflated in a typical experiment. This is the question about which there are divided opinions (Durgin, 2017; Schnall, 2017). So far, there are no strong arguments for the penetrability of vision since almost all behavioral findings taken in support of this idea can actually be attributed to the action of non-perceptual factors (judgment, memory, recognition, and allocation of attention) because of insufficient experimental control. Therefore, it remains an open question as to whether it is possible to design an experiment in which the direct influence of cognition and/or emotions on visual perception will be unambiguously isolated.

2.1.4. A Model of the Interaction Between Color Perception and Color Memory

Marić and Domijan (2018b) in their review paper suggested that there are sufficient grounds to defend the thesis that vision is cognitively impenetrable. However, it still remains an open question of how vision protects itself from top-down influences in the face of feedback connections in the visual hierarchy. Marić and Domijan (2020) adopted a theoretical framework of the adaptive resonance theory to show that it is possible to regulate and constraint the top-down influences in order to prevent them from altering the content of visual perception. To make the argument concrete, Marić and Domijan (2020) developed the color ART circuit to address the issue of how prior knowledge affects color perception. The model explains how traces of erroneous expectations about incoming color are eventually removed from the color perception, although their transient effect may be visible in behavioral responses or in brain imaging. This study demonstrates that the color ART circuit is specific computational implementation of the predictive coding system that is not penetrable by top-down influences (see Appendix D for the full paper and Appendix E for supplemental materials containing the MATLAB code to reproduce all results).

2.1.4.1. Introduction

The answer to the question of whether cognitive processes can directly alter the content of visual perception is not clear yet. One perspective holds that there is the complete independence of vision and cognition on the grounds that the computational role of vision is to provide the accurate representation of the environment. This is known as a cognitive impenetrability of visual perception (Pylyshyn, 1999; Raftopoulos, 2014). A contrary perspective proposes continuity of visual perception and cognition based on the idea that cognition provides contextual information that can resolve contradictory sensory evidence or can fill in missing parts. This position arises from predictive coding models (Clark, 2013; Friston, 2010; Hohwy, 2013, 2017).

To provide a clear demonstration of the fact that massive feedback connections in the visual cortex do not imply cognitive penetration of visual perception, Marić and Domijan (2020) adopted the adaptive resonance theory (ART) proposed by Grossberg and colleagues as a guiding theoretical framework. ART is an alternative theoretical framework that deals with predictions and top-down expectations differently compared to predictive coding models. ART’s mechanism of the novelty detection is responsible for protecting perception from top-down effects. They emphasized that the cognitive impenetrability of visual perception is a natural consequence of brain mechanisms designed to prevent interference between new inputs and old memories. Top-down signals carrying predictions can influence only early stages of processing in the ART circuit, that is, before resonance occurs. After the resonance is established, no further cognitive influence is possible because the resonant state represents conscious experience of recognizing a familiar input pattern. In addition, the ART circuit generates impenetrable percepts because it is an attractor network that generates phenomenon known as *hysteresis*. An attractor is a stable network state that is resistant to change. In other words, the network activity cannot leave this state once it is drawn to it, so the network is resistant to external perturbations. Hysteresis is a consequence of a strong excitatory loop between the F_1 and F_2 layer. This is also the reason why the ART network requires an external reset in the first place (Francis et al., 1994).

With respect to color vision, Marić and Domijan (2020) proposed that conscious experience of color arises from the resonance between color category (hue) encoded in F_2 layer and specific pattern of color-opponent activation registered in the F_0 layer. In other words, each discernible color requires its own separate F_2 node akin to a color grandmother cell (Bowers, 2009, 2017). Vigilance parameter that controls the match between bottom-up and top-down inputs has to be set to a high value in order to enable discrimination of a large number of hues. If prediction is disconfirmed by mismatching bottom-up activity, it cannot take any part in conscious perceptual experience by virtue of the orienting subsystem. Furthermore, Marić and Domijan (2020) argued that behavioral and neural data supporting cognitive penetrability of vision actually capture transients of neural activity that occurs during the processing of erroneous expectations. The ART circuit requires some time to complete its processing cycle, especially in the case of the mismatch computation. As a consequence, the ART response will reflect expectation rather than perception in the early stages of processing. This notion still does not support cognitive penetrability of visual perception because the ART circuit will eventually find the best possible hue category that matches with the bottom-up input. This will be signaled by the resonance between F_1 and F_2 layers, and resonance corresponds to our conscious experience of seeing a color (Grossberg, 2017).

There are numerous studies claiming to find support for CPV and it would be difficult to address all of them, so Marić and Domijan (2020) focused on color perception. Several findings suggest that knowledge about typical color associated with an object affects perceived color. Hansen et al. (2006) asked participants to adjust the color of natural or artificial objects to neutral gray. They found that achromatic adjustments of objects that are typically associated with one so called intrinsic color (e.g., tomato, broccoli, banana) were biased toward complementary colors. For example, participants adjusted banana to be slightly bluish. This is known as a memory color effect (Olkkonen et al., 2008, Witzel, 2016). It is explained as a consequence of the retrieval of object’s typical color from memory. Retrieval further creates a weak perceptual experience that forces participants to choose complementary color in order to offset the effect of retrieved intrinsic color. Lupyan (2015b) provided another compelling evidence for the cognitive penetrability of color perception. He reported that adapting to objects with intrinsic color (e.g., tomato) creates stronger afterimages (more vivid color) than adapting to arbitrarily colored objects (e.g., car). When used on natural images, this effect is known as the Spanish castle illusion. There is also functional neuroimaging evidence that it is possible to decode memory color from V4 activity when participant view achromatic image of the object with intrinsic color (Bannert & Bartels, 2013). Marić and Domijan (2020) showed that it is possible to simulate the memory color effect and Spanish castle illusion within the ART circuit, that is, within a computational architecture designed to protect perception from cognitive influences.

The model consists of four components: color pre-processing, the color ART circuit, feedback projections from the inter-ART associative map, and color working memory. In the first stage, color pre-processing, we simulated several processes that take place before the color ART circuit starts to encode hues. Marić and Domijan (2020) assumed that this stage computes the ratio between the activation of the single cone and the total cone output as in L/(L+M) or M/(L+M). This is achieved by divisive inhibition, which has several advantages (Foster, 2011; Hansen & Gegenfurtner, 2009; Hong & Tong, 2017; Seymour et al., 2016; Smithson, 2005) that are discussed in detail in Appendix D. We focused on the red-green opponency only to simplify the model. To account for the Spanish castle illusion, we needed an additional processing component capable of generating afterimages. This is achieved by incorporating transmitter habituation or synaptic depression in cone opponent channels. It refers to a reduction in the amount of available transmitter in response to sustained stimulation of presynaptic sites. When the transmitter habituation is embedded into a competitive circuit with two opponent pathways, it acts as a gate that shifts the competitive balance toward the unhabituated pathway (Francis, 2010; Grossberg, 1980).

The color ART circuit was based on the real-time implementation of the fuzzy ART algorithm (Carpenter et al., 1991). Pre-processing described in the previous paragraph leads to the activation of the F_0 layer. The F_1 layer computes a fuzzy intersection between the bottom-up input from the F_0 layer and read-out of top-down adaptive weights arriving from the F_2 layer. Layers F_0 and F_1 consist of cone-specific nodes, and the F_2 layer consists of nodes encoding color categories, i.e., hue cells. Hue cells selectively respond to the pattern of activation across L and M opponent input occurring across F_1 nodes. The F_2 layer is modeled as a WTA network that chooses node receiving maximal input from the F_1 layer and inhibits all other nodes. In this way, the F_2 layer transforms the linear response of cone-opponent input found in V1 into a non-linear color-tuned response found in the globs of the posterior IT (Bohon et al., 2016; Conway, 2009; Conway et al., 2007; Zaidi et al., 2014). When the selected hue node in the F_2 layer establishes resonance with the F_1 layer, it represents the perceived hue. The hue cells also receive feedback projections from the inter-ART associative map, which is going to be described in the next paragraph.

To explain how object knowledge affects color perception we followed the same approach previously employed by Domijan and Šetić (2016). We proposed that there exist two parallel ART circuits in the ventral stream; one is dedicated to color coding (the color ART circuit), and another is dedicated to shape coding (the shape ART circuit). This reflects the division of labor between color-selective and form-selective neurons found in the ventral visual stream (Lennie, 1998). We assumed that F_2 layers of the both color and shape ART circuits are mutually connected via the inter-ART associative map. Synaptic connections between winning nodes in the color and shape ART circuits and the inter-ART associative map are established via Hebbian learning. Such combined selectivity to color and form has been observed in the anterior IT (Chang et al., 2017; Lafer-Sousa & Conway, 2013). The inter-ART associative map encodes statistical regularities of co-occurrence of colors and shapes. In other words, the inter-ART associative map establishes strong bi-directional associations between those combinations of colors and shapes that often occur together in the environment (e.g., yellow ↔ banana). Next, we followed O'Callaghan et al. (2017) and suggested that the shape ART circuit responds faster to the same stimulus relative to the color ART circuit. This implies that the shape ART circuit may activate inter-ART associative map which further sends top-down expectations about color to the color ART circuit. Such top-down activation of the F_2 layer biases processing within the color ART circuit when it receives its bottom-up input. In the simulations reported in the paper, the shape ART circuit and inter-ART associative map were not explicitly modeled because they served here just to explain how feedback signals to the color ART circuit were generated.

With respect to the anatomical location in the brain, we hypothesized that the color ART circuit resides in V4 or it is spanned between V4 and the posterior IT. In the latter case, V4 may encompass the F_0 and F_1 layer, and the posterior IT the F_2 layer (Papale et al., 2018; Roe et al., 2012; Winawer & Witthoft, 2015). Furthermore, we assumed that color categories are learned from spatially invariant representation that receives converging input from all V1 locations. Which location will send activity to the color ART circuit depends on spatial attention that dynamically routes signal flow from V1 via V2 to V4. The model implies that only a small number of colors can consciously be perceived arriving from the retinotopic locations to which spatial attention is directed (Duncan, 1980a, 1980b; Huang & Pashler, 2007; Huang et al., 2007).

In addition to the bias induced by the inter-ART associative map on the activity of hue cells in the color ART circuit, there is also possibility that some of the effects of expectations arises from the transfer of color signals to color working memory (Beck & Schneider, 2017). To account for this case, we added color working memory circuit to the model. It receives input from the color ART circuit. We assumed that the color working memory circuit is also a WTA network like the F_2 layer of the color ART circuit. We further assumed that the color working memory circuit contains the identical hue representation as the one found in the color ART circuit. To account for variability in color reports and color confusions observed in the color working memory, the input to the color working memory circuit is passed through a distance-dependent filter before it activates competition in the color working memory circuit. In this way, we modeled how working memory and/or decision-making circuits may contribute to the biased reports of color perception as observed in behavioral studies.

2.1.4.2. Simulation of the Memory Color Effect

Before the simulation of the memory color effect, we performed the simulation of a stable self-organization of hue categories to illustrate that the color ART circuit is able to learn hue categories without catastrophic forgetting. This simulation illustrated the stability of learning in the color ART circuit. Input to the cones ranged from 0 to 10 and it was systematically increased or decreased in small steps. The F_2 layer consisted of 21 nodes labeled as Node 1 through Node 21. At the end of the learning session, each F_2 node was tuned to one hue ranging from pure red (Node 1 was tuned to input I_L = 10.0, I_M = 0.0) across various mixtures of red and green (Nodes between 2 and 20) to pure green (Node 21 was tuned to input I_L = 10.0, I_M = 0.0). The color tuning established in this simulation was used in all subsequent simulations reported in this study.

Next, we simulated the memory color effect by employing gray as the input to the color ART circuit. The bottom-up input was set at I_L = 5.0, I_M = 5.0 from t = 200 ms until the end of the simulation. Another input to the F_2 layer came from top-down color expectation from the inter-ART map. We assumed that a single node in the inter-ART map is connected to multiple nodes in the color ART circuit because we can observe the same object with different colors in different occasions. In subsequent simulations we examined the impact of two hypothetical distributions on the activity of F_2 nodes, which we labeled as *narrow* and *wide feedback*. The narrow distribution corresponds to a situation where the observer had a narrow range of color experiences associated with an object, and the wide distribution corresponds to a situation where color is less diagnostic because the observer might have more varied color experiences associated with an object.

First, we examined the activity of opponent L- and M-cone pathways in response to the presentation of gray color. The simulation showed that the activity of cones tracked input amplitudes. Transmitter gates faithfully carried over input amplitudes from the cones to the F_0 nodes. The F_0 nodes exhibited normalization of cone output due to divisive inhibition. After an initial burst of activity due to disynaptic inhibition, both F_0 nodes settled to 0.5 thus illustrating their ability to represent relative contributions of the L- and M- cone output. The F_1 nodes successfully tracked the activity of F_0 nodes, except during a short period of time from about 300 ms to 400 ms. This activity reduction was a consequence of the activation of the F_2 node at location 3 that was the winner of the competition due to the presence of narrow feedback with gain factor 0.1. We assumed that the arrival of the bottom-up input was delayed relative to the arrival of feedback signals to the F_2 layer. Instead of the winner was the Node 11, the winning node was shifted in the direction of feedback to Node 3 that consisted of strong L-cone (I_L = 9.0) and weak M-cone (I_M = 1.0) activation. Consequently, the activity reduction in the M-cone pathway arise from the weak top-down signal.

Second, we examined how the F_2 layer handles the influence of narrow and wide feedback at three levels of the feedback gain factor 0.1, 0.3, and 0.5. Results showed that the gain factor of 0.3 was a reasonable upper limit in both types of the feedback distributions because further increase in the feedback strength would result in an exhaustive search across all F_2 nodes, that is, of all nodes positioned at locations from 1 to 10. Therefore, this setting of the gain factor was used in all subsequent simulations. With respect to cognitive penetrability of color perception, we observed that in each instance examined the F_2 layer eventually reached the Node 11, which corresponds to the exact mid-point between red and green. However, the time needed to select this node varied as a function of the strength and shape of the feedback distribution. Generally, widening the spread of feedback distribution as well as increasing the feedback gain factor enabled the network to spend more time in a biased state, thus increasing the chance to detect the memory color effect. The simulation showed that moving the peak of the feedback distribution toward Node 1 or Node 21 makes the memory color effect stronger and moving the peak closer to Node 11 makes the effect weaker.

Third, we showed how the timing of arrival of feedback signals relative to the bottom-up input influenced the activity of the F_2 layer. Feedback was turned on at t = 250 ms, t = 275 ms, or t = 300 ms, thus creating feedback delays of 50 ms, 75 ms, or 100 ms, respectively. Only when the feedback was delayed for 50 ms the F_2 layer was biased toward red hues to the same degree as observed when there was no delay at all in both narrow and wide feedback. Results indicate that there is a narrow temporal window during which feedback signals are able to penetrate the F_2 layer and bias color perception toward diagnostic hues. In other words, the amount of time needed to reach resonance depends on the relative timing of feedback signals and the bottom-up input.

Fourth, we showed how the speed of neural activation (a node's time constant τ_x) influenced the penetrability of the F_2 layer. We varied the time constant of the F_2 node between 100 ms, 50 ms, and 25 ms to make it slower relative to the F_1 layer and other model components. When the speed of neural activity was set at τ_x = 100 ms and τ_x = 50 ms, the memory color effect completely disappeared in the condition of narrow feedback, while in the condition of wide feedback produced weak but long lasting (at τ_x = 100 ms), and stronger but short-lasting (at τ_x = 50 ms) memory color effect. When the speed of neural activity was set to τ_x = 25 ms, both narrow and wide feedback produced the memory color effect. Therefore, it can be concluded that the speed of neural integration within nodes in the F_1 and F_2 layers may also contribute to the total amount of time the network will spend in the hysteretic state that is not supported by the bottom-up input. If the observer chooses to respond prior to the occurrence of the reset signal, then his report will reflect bias induced by the top-down expectation. However, if the observer waits for a while until the dynamics of the ART circuit settle on the unbiased, long-lasting resonant state, then no memory color effect would occur.

Finally, we examined the behavior of the F_2 layer under different choices of vigilance parameter ρ. Vigilance was set at 0.5, 0.7, or 0.9 in the interval [0 ms, 800 ms]. Results showed that vigilance parameter contributes to a great variability in the strength of a memory color effect. Low level of vigilance such as 0.5 enabled both narrow and wide feedback to penetrate the F_2 layer and to generate the resonant state with the feedback-biased choice of hue. When vigilance was increased to 0.7, only wide feedback remained in a biased state. After vigilance was returned to its default value of 0.96 at t > 800 ms, in all conditions examined, the F_2 layer inhibited the current winner and selected Node 11, unless it was not already selected before. So, even if the ART circuit spends some time in a state of low vigilance where resonance is established with a biased hue category, it will eventually recover from such condition and reinstate its sharp distinction between hue categories.

2.1.4.3. Simulation of the Spanish Castle Illusion

To simulate the Spanish castle illusion, we compared two situations. The first simulation showed how color pre-processing generates a typical afterimage without the presence of feedback signals, while the second simulation showed how feedback from the inter-ART associative map to the F_2 layer makes color sensation stronger (more vivid) in the feedback-enhanced afterimage.

The network was first adapted to red color, and then gray color was presented. Thus, the bottom-up input was set to $I_L = 10.0$ and $I_M = 0.0$ in the interval between $t = 100$ ms and $t = 600$ ms, and then to $I_L = 5.0$ and $I_M = 5.0$ until the end of the simulation. During the maintenance of an elevated activity level in the L-cone pathway, its neurotransmitter gate decayed and became less effective because of the exhaustion of its presynaptic buttons. In the F_0 layer, there was an initial overshoot in the L-cone pathway in response to the presentation of red color, which was followed by a small dip in its activity because of transmitter depletion. After the network was adapted to red, the subsequent presentation of mid-gray stimulus activated both cone pathways to the same degree. However, the response of the F_0 node in the M-cone pathway was stronger than that of the L-cone because of the imbalance in the amount of neurotransmitter available in two pathways. As a consequence of this imbalance, there was an activity overshoot in the M-cone pathway. The temporal evolution of the activity of F_1 nodes follows a similar pattern as observed in the respective F_0 nodes. Their activity level was generally lower relative to the F_0 nodes because they compute the fuzzy intersection. The activity of the F_2 layer and the orienting subsystem tracked changes that occurred in the F_1 layer. At the beginning, the F_2 layer selected Node 1 as a winner, indicating perception of pure red. When the input color switched to gray, activity overshoot in the F_1 node of the M-cone pathway triggered activation of the orienting subsystem and forced selection of Node 15 in the F_2 layer, which corresponds to perception of the slightly greenish hue and represents bias toward green.

The second simulation demonstrated the Spanish castle illusion. Feedback was delivered to the F_2 layer from $t = 600$ ms until the end of simulation. We varied the orientation of feedback distribution in order to induce bias to the red or green hues. Also, we compared the effect of narrow and wide feedback as in the previous simulations. When the red-biased narrow feedback was applied, the simulation showed results consistent with the experimental report of Lupyan (2015b) who observed no statistically significant difference between conditions where bias was induced toward adapted hue and control condition without feedback. On the other hand, the green-biased narrow feedback forced the F_2 layer to choose Node 18 that corresponds to a more vivid experience of green compared to Node 15 chosen in control condition without feedback. Thus, the narrow feedback increased vividness of the experience of green color in the afterimage as observed in the Spanish castle illusion. However, this feedback-enhanced afterimage is fleeting and is removed at around 1,100 ms, which explains why we eventually arrive at the realization that we are watching an achromatic scene. In contrast to the narrow feedback, the wide feedback failed to produce the Spanish castle illusion. The red-biased wide feedback actually produced weaker color experience than normal afterimage would, and the green-biased wide feedback could not generate feedback-enhanced afterimage at all.

2.1.4.5. Conclusion

The ART circuit is an attractor neural network equipped with the computational mechanisms such as gain control and orienting subsystem that successfully constrain the impact of top-down predictions on the ongoing neural processing. Their computational role is to solve the stability-plasticity dilemma. This is achieved by preventing recoding of old memories in the face of new input patterns. Marić and Domijan (2020) found that the ART circuit is temporarily vulnerable to top-down influences during the period of searching through the state space to find the best match between the bottom-up input and choice of hue category. After the network reaches the attractor, it is fully resistant to any further top-down influence. Based on these observations, it can be concluded that the ART circuit is an example of the predictive system that is not penetrable (most of the time) by top-down influences. Taken together, reported findings suggest that the second problem of the thesis is successfully addressed.

2.2. Integrated Nature of Findings

Feedback projections from higher-order to lower-order areas are prominent feature of the visual cortex. Despite great efforts, there is still no consensus on what the role of feedback projections in visual perception is and how they affect our perceptual experience. In the thesis, computational modeling is employed to examine two contrasting hypotheses. Feedback may direct visual attention to a relevant portion of visual space, thus selectively routing signal flow through visual hierarchy. Conversely, in line with predictive coding framework, feedback may communicate expectations to lower-order areas in the visual hierarchy. Predictions are derived from internal or generative models of environment that encapsulate the knowledge of what is likely interpretation of incoming sensory stimulation. On this view, predictions directly modulate visual perception. In the thesis, neural network models were developed with the aim to show that the effect of predictions on visual perception is constrained by the same mechanisms that assure stable learning and memory as suggested by the adaptive resonance theory.

The first part of the thesis that encompasses work published in Marić and Domijan (2018a) and Marić and Domijan (2019) dealt with the proposal that feedback projections contribute to attentional selection of relevant information. Authors developed a new version of WTA network that was labeled the feature-based WTA or F-WTA network. The F-WTA network goes beyond previous proposals for spatial selection because it can select all locations occupied by the same feature value. In this way, it implements Boolean map theory of visual attention. Importantly, its capability to select locations does not depend on the number of locations sharing the same feature value to be selected. The F-WTA network is also capable of selecting objects and to simulate properties of mental contour tracing. To achieve this, the network is equipped with the mechanisms of dendritic and synaptic computation. Marić and Domijan showed that they are essential to achieve the flexibility required to respond to different types of top-down demands.

The second part of the thesis that encompasses work published in Marić and Domijan (2018b) and Marić and Domijan (2020) addressed the question of what the consequences for visual perception are if feedback transmits top-down predictions. In contrast to predictive coding, they suggested that feedback can be constrained by neural mechanisms used to assure stable memory and learning as shown by the adaptive resonance theory. ART has been successfully applied in modeling a wide range of behavioral and neural data. Here, they adopted this theory to develop a neural network for color perception and employed it to simulate effects such as memory color effect of Spanish castle illusion. These effects were taken as a support for the claim that predictions penetrate visual perception and consequently alter its content via feedback connections within the ventral stream. In contrast, their simulations showed that the same effects could be observed within the neural model that actively opposes feedback influences when they mismatch with the bottom-up input. They concluded that their simulations provide further support for the validity of ART as a theory of cognitive and neural information processing.

Taken together, results of the doctoral thesis suggest that the effects of feedback on visual processing are complex and diverse and that they cannot be reduced to a single process. The feedback can produce different outcomes depending on the type of neural architecture in which it is embedded. Within the dorsal stream feedback contributes to attentional selection by providing top-down guidance to the spatial map that flexibly selects locations, objects, or features. In contrast, within the ventral stream feedback participates in a formation of the resonant state that corresponds to our conscious perceptual experience. In this network, feedback is controlled by bottom-up inputs, not the other way around. Consequently, feedback within the ART circuit cannot alter the content of perception. This conclusion runs contrary to currently popular predictive coding framework that assumes pervasive effect of feedback at all levels of visual hierarchy.

2.3. Scientific Contribution of the Thesis

With respect to the modelling of visual selective attention, the doctoral thesis showed that the WTA neural network endowed with mechanisms of synaptic and dendritic computation can simultaneously support space-, object-, and feature-based selection. In this way, the proposed model provides a unified explanation of seemingly disparate attentional processes and goes far beyond previous WTA models that are restricted to modeling space-based attention. It was also shown that the same model is capable of implementing visual routines such as mental contour tracing. Proposed model provides a neural interface for the interaction between visual perception and cognition. Previous theoretical work treated dendritic and synaptic computation in isolation and independent from computational demands of attentional selection. In this work, these computational elements were brought together in a unique neural model to show that they greatly expand the capability of WTA network to flexibly select either location in space, or all locations occupied by the object, or all locations occupied by the same feature value.

With respect to the question of the role of predictions in visual perception, the doctoral thesis helps to clarify that the existence of feedback projections in the visual cortex do not necessarily imply that that top-down influences can alter the content of conscious visual perception. The ART computational framework is adopted to show that feedback projections are a necessary component of the neural architecture capable of stable learning and memory. Feedback supplies top-down expectations that are compared with bottom-up input. However, at the same time, feedback projections are prevented from the direct influence on visual perception by the same mechanisms that enable such stability.

3

CONCLUSIONS

3. CONCLUSIONS

3.1. Feedback and Visual Attention

Winner-takes-all networks are important class of neural models that are used in modeling visual selective attention. The doctoral thesis showed how to incorporate dendritic and synaptic computation to expand the flexibility of a standard WTA network. The new model, labeled as the F-WTA network, is capable of selecting multiple winners depending on top-down guidance provided by feedback connections. If only one location is cued, the network spreads the enhanced activity from the cued location to all connected elements. Results of computer simulations showed that the model's behavior is consistent with psychophysical findings on mental contour tracing suggesting that the proposed network successfully implements visual routines and object-based attention. On the other hand, if the value of an abstract feature, such as red color, is cued then the proposed network selects all locations occupied by red color, thus acting as a Boolean map. Moreover, computer simulations showed that the network can combine multiple Boolean maps using set operations of intersection and union. Taken together, the results showed that the F-WTA network unified Boolean map theory of visual attention (Huang & Pashler, 2007) with incremental grouping theory (Roelfsema & Houtkamp, 2011) to provide a more complete theory of visuospatial selection.

3.2. Feedback and Predictions

According to the predictive coding framework, feedback connections send predictions to the lower-order stages of cortical visual processing. Predictions embody knowledge about the likely causes of stimulation registered in the retina. Predictive coding implies that visual perception is continuous with cognition, that is, there is no clear boundary between them. Accordingly, there is much behavioral and neural data to support the conclusion that vision is cognitively penetrable. The doctoral thesis showed that ART is a viable neurocomputational alternative to predictive coding. ART handles feedback influences on vision in a different way. In ART, the mismatch between bottom-up input and top-down predictions causes an activation of the orienting subsystem that releases an inhibitory wave to reset the network and clear off traces of erroneous prediction. A new version of ART network called the color ART circuit is developed and designed to simulate properties of color perception. Computer simulations of the color ART circuit showed that the same empirical data that seem to support cognitive penetration, such as memory color effect of Spanish castle illusion, are also observed within the color ART circuit. It was concluded that feedback projections may communicate predictions, which does not imply, however, that vision is cognitively penetrable.

4

REFERENCES

REFERENCES

Abbott, L. F., & Regehr, W. G. (2004). Synaptic computation. *Nature, 431*(7010), 796–803. https://doi.org/10.1038/nature03010

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences,* 8(10), 457–464. https://doi.org/fh8cch

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current opinion in neurobiology, 46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Alger, B. E. (2002). Retrograde signaling in the regulation of synaptic transmission: Focus on endocannabinoids. *Progress in Neurobiology, 68*(4), 247–286. https://doi.org/10.1016/S0301-0082(02)00080-1

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13), 14.1–10. https://doi.org/10.1167/7.13.14

Anstis, S. (2002). Was El Greco astigmatic? *Leonardo, 35*, 208–208. https://doi.org/cr6ft6

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience, 10*(10), 1313–1321. https://doi.org/10.1038/nn1975

Ashby, F. G., & Hélie, S. (2011). A tutorial on computational cognitive neuroscience: Modeling the neurodynamics of cognition. *Journal of Mathematical Psychology, 55*, 273–289. https://doi.org/10.1016/j.jmp.2011.04.003

Banerjee, P., Chatterjee, P., i Sinha, J. (2012). Is it light or dark? Recalling moral behavior changes perception of brightness. *Psychological Science, 23*, 407–409. https://doi.org/10.1177/0956797611432497

Bannert, M. M., & Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology, 23*(22), 2268–2272. https://doi.org/10.1016/j.cub.2013.09.016

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.–M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., & Fries, P. (2015). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. *Neuron*, *85*(2), 390–401. https://doi.org/10.1016/j.neuron.2014.12.018

Beck, J., & Schneider, K. (2017). Attention and mental primer. *Mind & Language, 32*(4), 463–494. https://doi.org/10.1111/mila.12148

Bell, A. H., Pasternak, T., & Ungerleider, L. G. (2014). Ventral and dorsal cortical processing streams. In J. S. Werner and L. M. Chalupa (Eds.), *The new visual neurosciences* (pp 227–242). MIT Press.

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 1076–1096. https://doi.org/10.1037/0096-1523.25.4.1076

Bickford, M. E. (2016). Thalamic circuit diversity: Modulation of the driver/modulator framework. *Frontiers in Neural Circuits, 9*, 86. https://doi.org/10.3389/fncir.2015.00086

Bogler, C., Bode, S., & Haynes, J. D. (2011). Decoding successive computational stages of saliency processing. *Current Biology, 21*(19), 1667–1671. https://doi.org/10.1016/j.cub.2011.08.039

Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 complex). *eNeuro, 3*(4). https://doi.org/10.1523/ENEURO.0039-16.2016

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review, 116*, 220–251. https://doi.org/10.1037/a0014462

Bowers, J. S. (2017). Grandmother cells and localist representations: A review of current thinking. *Language, Cognition and Neuroscience, 32*, 257–273. https://doi.org/10.1080/23273798.2016.1267782

Branco, T., & Häusser, M. (2010). The single dendritic branch as a fundamental functional unit in the nervous system. *Current Opinion in Neurobiology*, *20*(4), 494–502. https://doi.org/10.1016/j.conb.2010.07.009

Briggs F. (2020). Role of feedback connections in central visual processing. *Annual Review of Vision Science, 6*, 313–334. https://doi.org/10.1146/annurev-vision-121219-081716

Brooks, J. L. (2014). Traditional and new principles of perceptual grouping. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 57–87). Oxford University Press.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*, 123–152.

Callaway, E. M. (2005). Structure and function of parallel pathways in the primate early visual system. *Journal of Physiology, 566*, 13–19. https://doi.org/drkq3p

Carpenter, G. A., & Grossberg, S. (Eds.). (1991). *Pattern recognition by self-organizing neural networks.* MIT Press. https://doi.org/10.7551/mitpress/5271.001.0001

Carpenter, G. A., & Grossberg, S. (2003). Adaptive resonance theory. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks, Second Edition* (pp. 87–90). MIT Press.

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks, 4*(6), 759–771. https://doi.org/10.1016/0893-6080(91)90056-B

Chang, L., Bao, P., & Tsao, D. Y. (2017). The representation of colored objects in macaque color patches. *Nature Communications, 8*(1), 2064. https://doi.org/10.1038/s41467-017-01912-7

Cheadle, S., Egner, T., Wyart, V., Wu, C., & Summerfield, C. (2015). Feature expectation heightens visual sensitivity during fine orientation discrimination. *Journal of Vision,* 15(14), 14. https://doi.org/10.1167/15.14.14

Chen Z. (2012). Object-based attention: A tutorial review. *Attention, perception & psychophysics*, *74*(5), 784–802. https://doi.org/10.3758/s13414-012-0322-z

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 1–73. https://doi.org/f4xkv5

Cole, S., Balcetis, E., i Dunning, D. (2013). Affective signals of threat increase perceived proximity. *Psychological Science*, *24*(1), 34–40. https://doi.org/f6t5

Conway, B. R. (2009). Color vision, cones, and color-coding in the cortex. *Neuroscientist, 15*(3), 274–290. https://doi.org/10.1177/1073858408331369

Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized color modules in macaque extrastriate cortex. *Neuron*, *56*(3), 560–573. https://doi.org/dmrm75

Crundall, D., Cole, G. G., & Underwood, G. (2008). Attentional and automatic processes in line tracing: Is tracing obligatory? *Perception & Psychophysics*, *70*(3), 422–430. https://doi.org/10.3758/PP.70.3.422

Davis, G., Driver, J., Pavani, F., & Shepherd, A. (2000). Reappraising the apparent costs of attending to two separate visual objects. *Vision Research, 40*(10–12), 1323–1332. https://doi.org/10.1016/S0042-6989(99)00189-3

Davis, G., Welch, V. L., Holmes, A., & Shepherd, A. (2001). Can attention select only a fixed number of objects at a time? *Perception, 30*(10), 1227–1248. https://doi.org/10.1068/p3133

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research, 20*(10), 847–856. https://doi.org/10.1016/0042-6989(80)90065-6

Domijan, D., & Šetić, M. (2016). Resonant dynamics of grounded cognition: Explanation of behavioral and neuroimaging data using the ART neural network. *Frontiers in Psychology*, *7*(139), 1–13. https://doi.org/10.3389/fpsyg.2016.00139

Drummond, L., & Shomstein, S. (2010). Object-based attention: Shifting or uncertainty? *Attention, Perception, & Psychophysics, 72(7),* 1743-1755. https://doi.org/10.3758/APP.72.7.1743

Duncan, J. (1980a). The demonstration of capacity limitation. *Cognitive Psychology, 12*(1), 75–96. https://doi.org/10.1016/0010-0285(80)90004-3

Duncan, J. (1980b). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87*(3), 272–300. https://doi.org/10.1037/0033-295X.87.3.272

Durgin, F. H., Baird, J., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review, 16*(5), 964–969. https://doi.org/10.3758/PBR.16.5.964

Edwards, G., Vetter, P., McGruer, F., Petro, L. S., & Muckli, L. (2017). Predictive feedback to V1 dynamically updates with sensory input. *Scientific Reports, 7*:16538. https://doi.org/10.1038/s41598-017-16093-y

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/bf03211502

Ermentrout, B. G. (1992). Complex dynamics in winner-take-all neural nets with slow inhibition. *Neural Networks*, *5*(3), 415-431.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. MIT Press.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Firestone, C. (2013). On the origin and status of the "El Greco fallacy". *Perception, 42*(6), 672–674. https://doi.org/10.1068/p7488

Firestone, C., & Scholl, B. J. (2014). "Top-dow" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science, 25*(1), 38–46. https://doi.org/10.1177/0956797613485092

Firestone, C., i Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences, 39*. https://doi.org/10.1017/S0140525X15000965

Firestone, C., & Scholl, B. J. (2017). Seeing and thinking in studies of embodied "perception". *Perspectives on Psychological Science, 12*(2), 341–343. https://doi.org/gh4g

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. MIT Press.

Foster, D. H. (2011). Color constancy. *Vision Research, 51*(7), 674–700. https://doi.org/10.1016/j.visres.2010.09.006

Francis, G. (2010). Modeling filling-in of afterimages. *Attention, Perception, & Psychophysics, 72*(1), 19–22. https://doi.org/10.3758/app.72.1.19

Francis, G. (2019). *Hypothesis testing reconsidered* (Elements in perception). Cambridge University Press. https://doi.org/10.1017/9781108582995

Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research, 34*, 1089–1104. https://doi.org/10.1016/0042-6989(94)90012-4

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology, 4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gage, N. M., & Baars, B. J. (Eds.) (2018). *Fundamentals of cognitive neuroscience. A beginner's guide. (2nd ed.)*. Academic Press. https://doi.org/10.1016/C2014-0-03767-7

Gilbert, C. D., i Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience, 14*(5), 350–363. https://doi.org/10.1038/nrn3476

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., i Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review, 106*(4), 795–834. https://doi.org/10.1037/0033-295x.106.4.795

Goodale, M. A., i Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience, 15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52*, 217–257. https://doi.org/10.1002/sapm1973523213

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87*(1), 1–51. https://doi.org/10.1037/0033-295X.87.1.1

Grossberg, S. (1988) Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks, 1*, 17-61.

Grossberg, S. (2013). Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks, 37*, 1–47. http://dx.doi.org/10.1016/j.neunet.2012.09.017

Grossberg, S. (2017). Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Networks, 87*, 38–95. http://dx.doi.org/10.1016/j.neunet.2016.11.003

Grossberg, S., & Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception & Psychophysics, 43*(3), 241–277. https://doi.org/10.3758/bf03207869

Haarmann, H., & Usher, M. (2001). Maintenance of semantic information in capacity limited item short-term memory. *Psychonomic Bulletin & Review, 8*(3), 568–578. https://doi.org/10.3758/bf03196193

Hahnloser, R. L. (1998). On the piecewise analysis of networks of linear threshold neurons. *Neural Networks, 11*(4), 691–697. https://doi.org/10.1016/S0893-6080(98)00012-4

Hahnloser, R., Douglas, R. J., Mahowald, M., & Hepp, K. (1999). Feedback interactions between neuronal pointers and maps for attentional processing. *Nature Neuroscience, 2*, 746–752. https://doi.org/10.1038/11219

Hahnloser, R. H., Seung, H. S., & Slotine, J.-J. (2003). Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Computation, 15*(3), 621–638. https://doi.org/10.1162/089976603321192103

Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research, 44*, 501–521. https://doi.org/10.1016/j.visres.2003.09.033

Hansen, T., & Gegenfurtner, K. R. (2009). Independence of color and luminance edges in natural scenes. *Visual Neuroscience, 26*(1), 35–49. https://doi.org/czxtfh

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367–1368. https://doi.org/10.1038/nn1794

Häusser, M., & Mel, B. W. (2003). Dendrites: Bug or feature? *Current Opinion in Neurobiology, 13*(3), 372–383. https://doi.org/10.1016/S0959-4388(03)00075-8

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804. https://doi.org/10.1016/s0896-6273(02)01091-7

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hohwy, J. (2017). Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Consciousness and Cognition, 47,* 75–85. https://doi.org/10.1016/j.concog.2016.09.004

Hollingworth, A., Maxcey-Richard, A. M., & Vecera, S. P. (2012). The spatial distribution of attention within and across objects. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 135–151. https://doi.org/10.1037/a0024463

Hong, S. W., & Tong, F. (2017). Neural representation of form-contingent color filling-in in the early visual cortex. *Journal of Vision, 17*(13), 10. https://doi.org/10.1167/17.13.10

Horn, D., & Usher, M. (1990). Excitatory–inhibitory networks with dynamical thresholds. *International Journal of Neural Systems, 1*(3), 249–257. https://doi.org/10.1142/S0129065790000151

Houtkamp, R., Spekreijse, H., & Roelfsema, P. R. (2003). A gradual spread of attention. *Perception & Psychophysics, 65*(7), 1136–1144. https://doi.org/10.3758/bf03194840

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114*(3), 599–631. https://doi.org/10.1037/0033-295X.114.3.599

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, 317, 823–825. https://doi.org/10.1126/science.1143515

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*, 194–203. https://doi.org/10.1038/35058500

Jolicoeur, P., Ullman, S., & Mackay, M. (1991). Visual curve tracing properties. *Journal of Experimental Psychology: Human Perception and Performance, 17*(4), 997–1022. https://doi.org/10.1037/0096-1523.17.4.997

Kaski, S., & Kohonen, T. (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Networks*, *7*, 973–984. https://doi.org/10.1016/S0893-6080(05)80154-6

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron, 100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Klein, O., Doyen, S., Leys, C., Magalhaes de Saldanha da Gama, P. A., Miller, S., Questienne, L., i Cleeremans, A. (2012). Low hopes, high expectations: Expectancy effects and the replicability of behavioral experiments. *Perspectives on Psychological Science, 7*(6), 572–584. https://doi.org/10.1177/1745691612463704

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology, 232*(1), 149–162. https://doi.org/10.1113/jphysiol.1973.sp010261

Lafer-Sousa, R., & Conway, B. R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nature Neuroscience, 16*(12), 1870–1878. https://doi.org/10.1038/nn.3555

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*(1), 12–18. https://doi.org/10.1016/s1364-6613(02)00013-x

Lamme, V. A. F., i Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience, 23*(11), 571–579. https://doi.org/10.1016/s0166-2236(00)01657-x

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27*, 889–935. https://doi.org/10.1068/p270889

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General, 135*(4), 501–512. https://doi.org/10.1037/0096-3445.135.4.501

Levine, D. S. (2019). *Introduction to neural and cognitive modeling*. *3rd edition*. Routledge.

Liverence, B. M., & Franconeri, S. L. (2015). Resource limitations in visual cognition. In R. Scott and S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–13). John Wiley and Sons.

London, M., & Häusser, M. (2005). Dendritic computation. *Annual Review of Neuroscience, 28*(1), 503–532. https://doi.org/10.1146/annurev.neuro.28.061604.135703

Lupyan, G. (2015a). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology, 6*(4), 547–569. https://doi.org/10.1007/s13164-015-0253-4

Lupyan, G. (2015b). Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*. *161*, 117–130. https://doi.org/ghh8c7

Lupyan, G. (2017a). Changing what you see by changing what you know: The role of attention. *Frontiers in Psychology, 8*(553). https://doi.org/10.3389/fpsyg.2017.00553

Lupyan, G. (2017b). How reliable is perception? *Philosophical Topics, 45*(1), 81–106. https://doi.org/10.17605/OSF.IO/R7SJJ

Macknik, S. L., & Martinez-Conde, S. (2009). The role of feedback in visual attention and awareness. In M. S. Gazzaniga (Ed.), *The cognitive neuroscience* (pp. 1165–1179). MIT Press.

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America, 70*(11), 1297–1300. https://doi.org/10.1364/JOSA.70.001297

Marić, M., & Domijan, D. (2018a). A neurodynamic model of feature-based spatial selection. *Frontiers in psychology, 9*, 417. https://doi.org/10.3389/fpsyg.2018.00417

Marić, M., & Domijan, D. (2018b). Mogu li kognicija i emocije utjecati na vid? [Can cognition and emotions affect vision?]. *Psihologijske teme, 27*(2), 311–338. https://doi.org/10.31820/pt.27.2.9

Marić, M., & Domijan, D. (2019). Neural dynamics of spreading attentional labels in mental contour tracing. *Neural Networks, 119*, 113–138. https://doi.org/10.1016/j.neunet.2019.07.016

Marić, M., & Domijan, D. (2020). A neurodynamic model of the interaction between color perception and color memory. *Neural Networks, 129*, 222–248. https://doi.org/10.1016/j.neunet.2020.06.008

Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science, 342*(6158), 1238406. https://doi.org/10.1126/science.1238406

Marr, D. (1982). *Vision*. Freeman.

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology, 14*(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences, 29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

McCormick, P. A., & Jolicoeur, P. (1991). Predicting the shape of distance functions in curve tracing: Evidence for a zoom lens operator. *Memory & Cognition, 19*(5), 469–486. https://doi.org/10.3758/BF03199570

McCormick, P. A., & Jolicoeur, P. (1992). Capturing visual attention and the curve tracing operation. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 72–89. https://doi.org/10.1037/0096-1523.18.1.72

McCormick, P. A., & Jolicoeur, P. (1994). Manipulating the shape of distance effects in visual curve tracing: Further evidence for the zoom lens model. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 48*(1), 1–24. https://doi.org/10.1037/1196-1961.48.1.1

Mel, B. W. (2016). Towards a simplified model of an active dendritic tree. In G. Stuart, N. Spruston, & M. Häusser (Eds.), *Dendrites. Third edition* (pp. 465–486). Oxford University Press.

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J. M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron, 89*(2), 384–397. https://doi.org/10.1016/j.neuron.2015.12.018

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785. https://doi.org/10.1016/j.neuropsychologia.2007.10.005

Minsky, M. L., & Papert, S. A. (1988). *Perceptrons: Expanded edition*. MIT Press.

Nassi, J. J., & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nature Reviews Neuroscience, 10*(5), 360–372. https://doi.org/10.1038/nrn2619

Newen, A., & Vetter, P. (2017). Why cognitive penetration of our perceptual experience is still the most plausible account. *Consciousness and Cognition, 47*, 26–37. https://doi.org/10.1016/j.concog.2016.09.005

Nobre, A. C., & Kastner, S. (2014). *The Oxford handbook of attention*. Oxford University Press.

O'Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., Jr., i Bar, M. (2017). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition, 47*, 63–74. https://doi.org/10.1016/j.concog.2016.05.003

O'Grady, R. B., & Müller, H. J. (2000). Object-based selection operates on a grouped array of locations. *Perception and Psychophysics, 62*(8), 1655–1667. https://doi.org/10.3758/bf03212163

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision, 8*(5), 13.11–16. https://doi.org/10.1167/8.5.13

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* The MIT Press.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT Press.

Papale, P., Leo, A., Cecchetti, L., Handjaras, G., Kay, K. N., Pietrini, P., & Ricciardi, E. (2018). Foreground-background segmentation revealed during natural image viewing. *eNeuro, 5*(3). https://doi.org/10.1523/ENEURO.0075-18.2018

Pooresmaeili, A., Poort, J., Thiele, A., & Roelfsema, P. R. (2010). Separable codes for attention and luminance contrast in the primary visual cortex. *Journal of Neuroscience, 30*(38), 12701–12711. https://doi.org/10.1523/jneurosci.1388-10.2010

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*, 341–365.

Raftopoulos, A. (2009). *Cognition and perception: How do psychology and neural science inform philosophy?* MIT Press.

Raftopoulos, A. (2014). The cognitive impenetrability of the content of early vision is a necessary and sufficient condition for purely nonconceptual content. *Philosophical Psychology, 27*(5), 601–620. https://doi.org/10.1080/09515089.2012.729486

Raftopoulos, A. (2019). *Cognitive penetrability and the epistemic role of perception*, Palgrave Innovations in Philosophy. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-10445-0_1

Raftopoulos, A., & Zeimbekis, J. (2015). The cognitive penetrability of perception: An overview. In J. Zeimbekis & A. Raftopoulos (Eds.), *The Cognitive Penetrability of Perception: New Perspectives*. Oxford University Press.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87. https://doi.org/10.1038/4580

Regehr, W. G., Carey, M. R., & Best, A. R. (2009). Activity-dependent regulation of synapses by retrograde messengers. *Neuron*, *63*(2), 154–170. https://doi.org/10.1016/j.neuron.2009.06.021

Richard, A. M., Lee, H., & Vecera, S. P. (2008). Attentional spreading in object-based attention. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 842–853. https://doi.org/10.1037/0096-1523.34.4.842

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research, 179*(1), 3–20. https://doi.org/10.1016/0006-8993(79)90485-2

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., & Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron, 74*(1), 12–29. https://doi.org/10.1016/j.neuron.2012.03.011

Roelfsema, P. R. (2005). Elemental operations in vision. *Trends in Cognitive Sciences, 9*(5), 226–233. https://doi.org/10.1016/j.tics.2005.03.012

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience, 29,* 203–227. https://doi.org/10.1146/annurev.neuro.29.051605.112939

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics, 73*(8), 2542–2572. https://doi.org/d3m47z

Roelfsema, P. R., Houtkamp, R., & Korjoukov, I. (2010). Further evidence for the spread of attention during contour grouping: A reply to Crundall, Dewhurst, and Underwood (2008). *Attention, Perception, & Psychophysics, 72*(3), 849–862. https://doi.org/10.3758/APP.72.3.849

Roelfsema, P. R., Khayat, P. S., & Spekreijse, H. (2003). Subtask sequencing in the primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 100*(9), 5467–5472. https://doi.org/10.1073/pnas.0431051100

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*(6700), 376–381. https://doi.org/10.1038/26475

Roelfsema, P. R., & Singer, W. (1998). Detecting connectedness. *Cerebral Cortex, 8*(5), 385–396. https://doi.org/10.1093/cercor/8.5.385

Rutishauser, U., & Douglas, R. J. (2009). State-dependent computation using coupled recurrent networks. *Neural Computation*, *21*(2), 478–509. https://doi.org/10.1162/neco.2008.03-08-734

Rutishauser, U., Douglas, R. J., & Slotine, J. J. (2011). Collective stability of networks of winner-take-all circuits. *Neural Computation, 23*(3), 735–773. https://doi.org/d8s62g

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5*(7), 631–632. https://doi.org/10.1038/nn876

Saenz, M., Buracas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research, 43*(6), 629–637. https://doi.org/ff246v

Schnall, S. (2017). Social and contextual constraints on embodied perception. *Perspectives on Psychological Science*, *12*(2), 325–340. https://doi.org/10.1177/1745691616660199

Schnall, S., Zadra, J. R., i Proffitt, D. R. (2010). Direct evidence for the economy of action: Glucose and the perception of geographical slant. *Perception, 39*(4), 464–482. https://doi.org/10.1068/p6445

Scholte, H. S., Spekreijse, H., & Roelfsema, P. R. (2001). The spatial profile of visual attention in mental curve tracing. *Vision Research, 41,* 2569–2580. https://doi.org/10.1016/S0042-6989(01)00148-1

Scimeca, J. M., & Franconeri, S. L. (2015). Selecting and tracking multiple objects. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(2), 109–118. https://doi.org/10.1002/wcs.1328

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron, 55*(2)*,* 301–312. https://doi.org/fs78wc

Seymour, K. J., Williams, M. A., & Rich, A. N. (2016). The representation of color across the human visual cortex: Distinguishing chromatic signals contributing to object form versus surface color. *Cerebral Cortex, 26*(5), 1997–2005. https://doi.org/10.1093/cercor/bhv021

Shaughnessy, J. J., Zechmeister, E. B., i Zechmeister, J. S. (2012). *Research methods in psychology* (9th ed.). McGraw Hill.

Sherman, S. M., & Guillery, R. W. (1998). On the actions that one nerve cell can have on another: distinguishing “drivers” from “modulators”. *Proceedings of the National Academy of Sciences of the United States of America, 95*(12), 7121–7126. https://doi.org/10.1073/pnas.95.12.pnas7121

Sherman, S. M., and Guillery, R. W. (2006*). Exploring the thalamus and its role in cortical function* (2nd ed.). MIT Press.

Shipp S. (2016). Neural Elements for Predictive Coding. *Frontiers in psychology*, *7,* 1792. https://doi.org/10.3389/fpsyg.2016.01792

Shomstein, S., & Yantis, S. (2002). Object-based attention: Sensory modulation or priority setting? *Perception & Psychophysics*, *64,* 41–51. https://doi.org/10.3758/BF03194556

Shomstein, S., & Yantis, S. (2004). Configural and contextual prioritization in object-based attention. *Psychonomic Bulletin & Review*, *11*, 247–253. https://doi.org/bqj358

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 360*(1458), 1329–1346. https://doi.org/10.1098/rstb.2005.1633

Song, H., Vonasch, A. J., Meier, B. P., i Bargh, J. A. (2012). Brighten up: Smiles facilitate perceptual judgement of facial lightness. *Journal of Experimental Social Psychology, 48,* 450–452. https://doi.org/10.1016/j.jesp.2011.10.003

Spillmann, L. (2014). Receptive fields of visual neurons: The early years. *Perception, 43*(11), 1145–1176. https://doi.org/10.1068/p7721

Spillmann, L., Dresp-Langley, B., & Tseng, C. H. (2015). Beyond the classical receptive field: The effect of contextual stimuli. *Journal of Vision*, *15*(9), 7. https://doi.org/10.1167/15.9.7

Spratling M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision research*, *48*(12), 1391–1408. https://doi.org/10.1016/j.visres.2008.03.009

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *The Journal of Neuroscience*, *30*(9), 3531–3543. https://doi.org/bsk486

Staadt, R., Philipp, S. T., Cremers, J. L., Kornmeier, J., & Jancke, D. (2020). Perception of the difference between past and present stimulus: A rare orientation illusion may indicate incidental access to prediction error-like signals. *PloS One*, *15*(5), e0232349. https://doi.org/10.1371/journal.pone.0232349

Stefanucci, J. K., & Geuss, M. N. (2009). Big people, little world: The body influences size perception. *Perception*, *38*(12), 1782–1795. https://doi.org/10.1068/p6437

Tao, H. W., & Poo, M. (2001). Retrograde signaling at central synapses. *Proceedings of the National Academy of Sciences*, *98*(20), 11009–11015. https://doi.org/bcj9v4

Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1628), 20130055. https://doi.org/10.1098/rstb.2013.0055

Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Tong, F. (2018). Foundations of vision. In J. T. Serences and J. T. Wixted (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, Volume 2, Sensation, Perception, and Attention, 4th Edition* (pp. 1–62). Joh Wiley & Sons.

Ullman, S. (1984). Visual routines. *Cognition*, *18*(1–3), 97–159.

Ullman, S. (1996). *High-level vision*. MIT Press.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. W. J. Mansfield (Eds.), *Analysis of Visual Behavior*. MIT Press.

Usher, M., & Cohen, J. D. (1999). Short term memory and selection processes in a frontal-lobe model. In D. Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience* (pp. 78–91). Springer-Verlag.

Valenti, J. J., & Firestone, C. (2019). Finding the "odd one out": Memory color effects and the logic of appearance. *Cognition*, *191*, 103934. https://doi.org/gh4f

Vandenbroucke, A. R. E., Fahrenfort, J. J., Meuwese, J. D. I., Scholte, H. S., i Lamme, V. A. F. (2016). Prior knowledge about objects determines neural color representation in human visual cortex. *Cerebral Cortex, 26*(4), 1401–1408. https://doi.org/10.1093/cercor/bhu224

Vatterott, D. B., & Vecera, S. P. (2015). The attentional window configures to object and surface boundaries. *Visual Cognition, 23*(5), 561–576. https://doi.org/gh3q

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling. *Philosphical Transactions of the Royal Society of London. B Biological Sciences, 372*(1714). https://doi.org/gg3qxc

Vecera, S. P., & Farah, M. J. (1997). Is image segmentation a bottom-up or an interactive process? *Perception & Psychophysics, 59*, 1280–1296.

Vetter, P., i Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition, 27*, 62–75. https://doi.org/10.1016/j.concog.2014.04.007

Winawer, J., & Witthoft, N. (2015). Human V4 and ventral occipital retinotopic maps. *Visual Neuroscience, 32*, E020. https://doi.org/10.1017/s0952523815000176

Witt, J. K., Proffitt, D. R., i Epstein, W. (2004). Perceiving distance: A role of effort and intent. *Perception, 33*(5), 577–590. https://doi.org/10.1068/p5090

Witzel, C. (2016). An easy way to show memory color effects. *i-Perception, 7*(5), 2041669516663751. https://doi.org/10.1177/2041669516663751

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object knowledge modulates colour appearance. *i-Perception, 2*(1), 13–49. https://doi.org/10.1068/i0396

Zaidi, Q., Marshall, J., Thoen, H., & Conway, B. R. (2014). Evolution of neural computations: Mantis shrimp and human color decoding. *i-Perception, 5*(6), 492–496. https://doi.org/10.1068/i0662sas

Zeimbekis, J. (2013). Color and cognitive penetrability. *Philosophical Studies*, *165*, 167–175. https://doi.org/10.1007/s11098-012-9928-1

Zhou, H., Schafer, R. J., & Desimone, R. (2016). Pulvinar-cortex interactions in vision and attention. *Neuron*, *89*(1), 209–220. https://doi.org/10.1016/j.neuron.2015.11.034

Zilberter Y., Harkany, T., & Holmgren, C. D. (2005). Dendritic release of retrograde messengers controls synaptic transmission in local neocortical networks. *The Neuroscientist, 11*(4), 334–344. https://doi.org/10.1177/1073858405275827

5

LIST OF FIGURES

LIST OF FIGURES

6

APPENDIX

APPENDIX

Appendix section consists of five appendices. Appendix A, B, C, and D represent one of the four published papers, respectively. Appendix E contains supplemental materials in the form of Open Science Framework links that generate MATLAB code to reproduce all results reported in papers. Every paper has been produced in collaboration with mentor who designed the study and wrote the manuscript, and PhD student performed computer simulations and wrote the manuscript.

APPENDIX A

A Neurodynamic Model of Feature-Based Spatial Selection

Marić, M., & Domijan, D. (2018). A neurodynamic model of feature-based spatial selection. *Frontiers in psychology*, *9*, 417. https://doi.org/10.3389/fpsyg.2018.00417

ABSTRACT

Huang and Pashler (2007) suggested that feature-based attention creates a special form of spatial representation, which is termed a Boolean map. It partitions the visual scene into two distinct and complementary regions: selected and not selected. Here, we developed a model of a recurrent competitive network that is capable of state-dependent computation. It selects multiple winning locations based on a joint top-down cue. We augmented a model of the WTA circuit that is based on linear-threshold units with two computational elements: dendritic nonlinearity that acts on the excitatory units and activity-dependent modulation of synaptic transmission between excitatory and inhibitory units. Computer simulations showed that the proposed model could create a Boolean map in response to a featured cue and elaborate it using the logical operations of intersection and union. In addition, it was shown that in the absence of top-down guidance, the model is sensitive to bottom-up cues such as saliency and abrupt visual onset.

Keywords: Boolean map, feature-based attention, lateral inhibition, neural network, winner-take-all

1. INTRODUCTION

In the literature on visual attention, significant progress has been made in characterizing the principles of selection. Visual attention can be allocated flexibly to a circumscribed region of space, the whole object or feature dimensions such as color and orientation (Nobre & Kastner, 2014). Indeed, early work suggested that a restricted circular region of space is a representational format of attentional selection. Posner (1980) proposed that attention operates like a spotlight that highlights a single circular region of space with a fixed radius. All locations that fall inside the spotlight are selected, and everything outside is left out. An extension of this proposal, which is called the zoom-lens model, suggests that the spotlight of attention can change its radius depending on the spatial resolution that one wants to achieve (Eriksen & St. James, 1986). If high resolution is required, the spotlight can be narrowed to capture details in the selected region, whereas the radius of the spotlight can be widened when a lower resolution is sufficient.

Other studies point to an object as a unit of selection. Duncan (1984) showed that it is easier to report two attributes if they appear on the same object, relative to the scenario in which each attribute appears on a different object. This finding implies that the object is selected as a whole and has been replicated many times using different stimuli and behavioral paradigms (Scholl, 2001). This effect cannot be explained by spatial attention because objects were spatially superimposed, that is, they shared the same locations. More recently, it was shown that attention can also be allocated to a visual feature such as color or direction of motion independent of spatial location (Saenz et al., 2002, 2003). Single-unit recordings have shown that feature-based attention is accompanied by the global location-independent modulation of neural response in a range of areas in the visual cortex (Boynton, 2005; Maunsell & Treue, 2006). Attentional modulation was described as a multiplicative gain change that increases responses of neurons that are selective to attended feature values and decreases responses of neurons that are tuned to unattended feature values (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999).

Object-based attention, however, is not necessarily detached from spatial representation. There is behavioral and neurophysiological evidence that object-based attention involves selection of all spatial locations that are occupied by the same object. Specifically, it was suggested that attention selects a grouped array of locations (O'Grady & Müller, 2000). In other words, attention spreads from one spatial location along the shape of the object and highlights all locations that belong to the object (Richard et al., 2008; Vatterott & Vecera, 2015).

Neurophysiological studies showed that object-based selection is indeed achieved by the spreading of the enhanced firing rate along the shape of the object (Roelfsema, 2006; Roelfsema & de Lange, 2016).

In a similar way, feature-based attention might involve the selection of all locations that are occupied by the same feature value, as shown by Huang and Pashler (2007). They proposed that attention is limited because it may access only one feature value (e.g., red) per dimension (e.g., color) at any given moment. However, the accessed feature value is bound to space in parallel, without capacity limits. Feature-based attention is allocated in space via the formation of a binary or Boolean map. When a conscious decision is made to attend to a specific feature value, the Boolean map indicates all spatial locations that are occupied by the chosen feature value because they are labeled by a positive value (e.g., 1), while all other locations are labeled with zero. In each selection process, selected locations need not be contiguous in space, but they must share the same feature value. After a Boolean map is formed, it is possible to operate on its output by applying the set operations of intersection and union. Recent work suggests that a spatial representation, such as a Boolean map, might mediate perceptual grouping by similarity (Huang, 2015; Yu & Franconeri, 2015). Moreover, the idea has been recently applied successfully in the computer vision literature on developing algorithms for saliency detection (Qi et al., 2017; Zhang & Sclaroff, 2016).

Figure 1 illustrates a Boolean map that is formed in response to three different stimulus configurations and sequential application of two top-down feature cues. Figure 1A shows a simple stimulus that consists of red and green squares. An observer might attempt to isolate only red or only green items. To do so, a top-down cue should be supplied to the feature map that encodes the desired feature value. For example, when attention is directed to the red color, the top-down cue highlights all locations that are occupied by red squares. The Boolean map picks up on this feature cue and forms a spatial representation in which cued locations are labeled with 1 (white) and non-cued locations are labeled with 0 (black). In terms of a neural network, these labels correspond to the active (excited) and inactive (inhibited) states of the corresponding nodes in the network (Boolean Map – 1). Later, the observer might wish to switch to green color (Boolean Map – 2). Again, in a response to a new feature cue, the Boolean map now shows all locations that are occupied by green squares.

Figure 1B shows a typical stimulus that is used in visual search experiments. It consists of red and green horizontal and vertical bars. The task is to find a red horizontal bar. This is an example of a conjunction search task in which two feature dimensions should be combined to find the target object. According to Huang and Pashler (2007), the conjunction task is solved in two steps. In the first step, a Boolean map is formed by top-down cueing of red items, irrespective of their orientations. In the second step, only horizontal items are cued. However, since red items have already been selected, the second Boolean map will correspond to the intersection of red and horizontal items. There is only one item that satisfies these selection criteria: the target. In this way, visual search is substantially faster compared to the strategy of sequentially visiting each item by moving the attentional spotlight across the visual field. It is also possible to reverse the order of the applied feature cues. In the first step, horizontal items might be cued, and the intersection is formed by highlighting red items in the second step. Importantly, there is behavioral evidence that observers indeed implement such a *subset selection* strategy in conjunction search tasks (Egeth et al., 1984; Kaptein et al., 1995). Moreover, Huang and Pashler (2012) showed that the same strategy is used in the perception of spatial structure in a stimulus that is composed of multiple items that differ in several dimensions.

Figure 1C illustrates an example of the union of two Boolean maps. As in the previous example, the observer starts by cueing red items and creating a Boolean map that consists of a representation of their locations. In the second step, the observer wishes to combine red with horizontal items. Therefore, in the second step, one should cue horizontal items but simultaneously maintain locations of the remaining items in memory. The resulting new Boolean map now represents the locations of all red and all horizontal items that were found in the image. Computing with Boolean maps might not be restricted to only two steps, as Figure 1 suggests. It is possible to incorporate more feature dimensions, such as motion, texture, or size, that can also be engaged in creating Boolean maps that are more complicated.

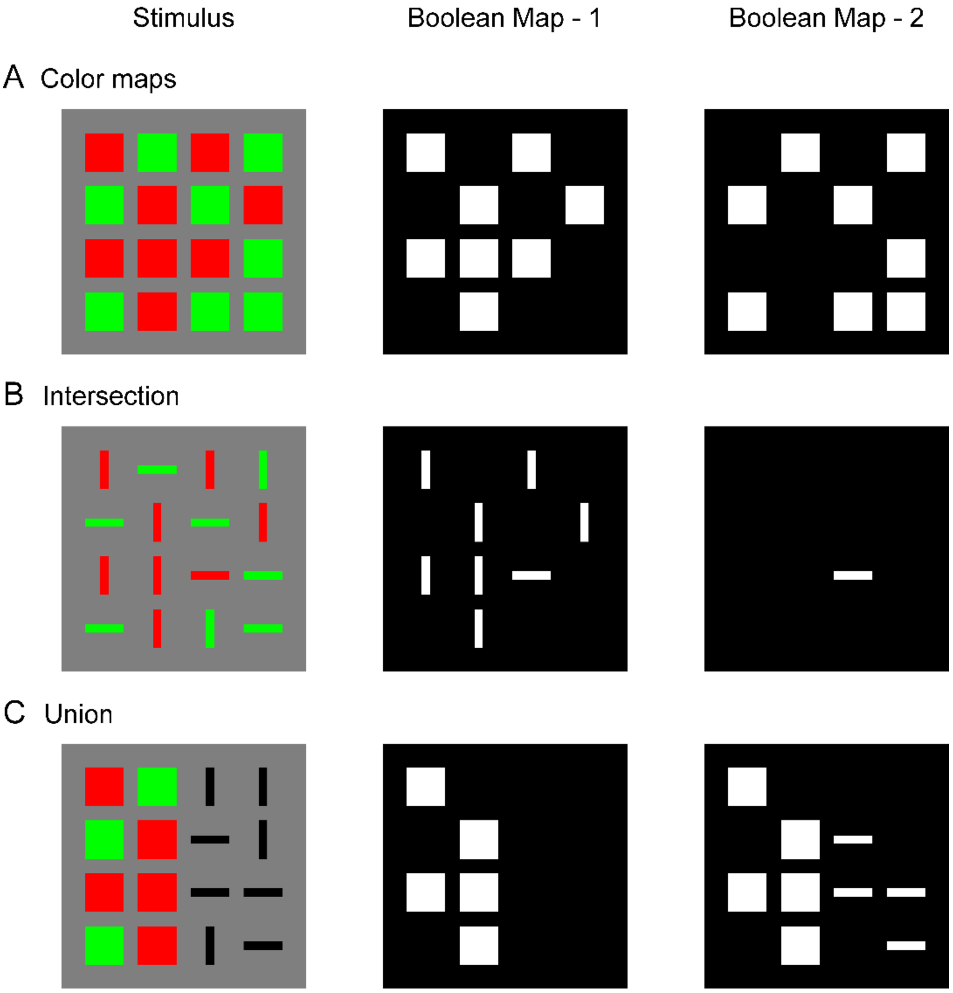


Figure 1. *Illustration of the Boolean map that was created in response to the input image (Stimulus) after the first feature cue was applied to the spatial representation (Boolean Map – 1) and after the second feature cue was applied (Boolean Map – 2). (A) Boolean maps that were created by two-color cues: red in the first step and green in the second step; (B) intersection of two Boolean maps, where red is cued in the first step and horizontal orientation in the second step; (C) union of two Boolean maps, where red is cued in the first step and horizontal orientation in the second step.*

Feature-based spatial selection, as illustrated by the Boolean map, provides a strong constraint on the computational models of visual attention because it requires simultaneous selection of arbitrarily many locations based on an arbitrary criterion that is set by the observer. Computational models of attention often rely on a winner-take-all (WTA) network to select a single, most salient location from the input image (Itti & Koch, 2000, 2001). The WTA network consists of an array of excitatory nodes that are connected reciprocally with inhibitory interneurons. This anatomical arrangement creates lateral inhibition among excitatory nodes that lead to the selection of a single node that receives maximal input and the suppression of all other nodes, which receive non-maximal input. However, when faced with the input where multiple (potentially many) nodes share the same maximal input level, the typical WTA network tends to suppress all winning nodes due to a strong mutual inhibition among them instead of selecting them together. For example, Usher and Cohen (1999) showed that, under the conditions of strong recurrent excitation and weak lateral inhibition, the WTA network reaches a steady state with multiple active winners. Importantly, activation of the winning nodes decreases linearly towards zero as their quantity increases. In other words, this network design suffers from the capacity limitation. This is a useful property in modeling short-term memory and frontal lobe function (Haarmann & Usher, 2001) but it is inadequate for understanding how the Boolean map might arise in a large retinotopic map, as exemplified by Figure 1.

Another problem is that the dynamics of the WTA network are not sensitive to transient changes in the input amplitude. Due to strong self-excitation and the resulting persistent activity, the WTA network settles into one of its memory states (fixed points). Importantly, each memory state is independent of later inputs. If self-excitation is weakened, the network will become sensitive to input. However, at the same time, it will lose its ability to form a memory state and will behave like a feedforward network (Rutishauser & Douglas, 2009). One way to solve this problem is to apply an external reset signal to the network before a new input is processed (Grossberg, 1980; Itti & Koch, 2000, 2001; Kaski & Kohonen, 1994). However, this is not sufficient in the context of feature-based attention. An intersection or union operation between two Boolean maps requires that the currently active memory state (formed after the first feature cue) be updated by taking into account new input (the second feature cue). Therefore, the dynamics of the WTA network should allow uninterrupted transition between memory states that are governed by external inputs. In other words, the WTA network should be capable of state-dependent computation (Rutishauser & Douglas, 2009).

To summarize, a WTA network that is capable of computing with Boolean maps should simultaneously satisfy two computational constraints:

1) It should be able to select together all locations that share a common feature value. This should be achieved without degrading the representation of the winners.
2) It should exhibit state-dependent computation, in which new inputs are combined with the current memory state to produce a new resultant state (e.g., intersection or union).

Here, we have developed a new WTA network that satisfies these constraints and provides the neural implementation of the Boolean map theory of attention (Huang & Pashler, 2007).

2. MODEL DESCRIPTION

The aim of the current work is to provide an explanation of how a Boolean map may be formed in a recurrent competitive network that can implement feature-based winner-take-all (F-WTA) selection. To this end, we have extended the previously proposed network model based on the linear-threshold units (Hahnloser, 1998; Hahnloser et al., 2003; Rutishauser & Douglas, 2009). Concretely, the model circuit is presented in Figure 2. It consists of a single inhibitory unit, which is reciprocally connected to a group of excitatory units. In addition to these basic elements, we introduce two processing components into the WTA circuit to expand its computational power. The first is a dendritic nonlinearity, which prevents excessive excitation that arises from self-recurrent and nearest-neighbor collaterals. We modeled the dendritic tree as a separate electrical compartment with its own non-linear output that is supplied to the node's body (Branco & Häusser, 2010; Häusser & Mel, 2003; London & Häusser, 2005; Mel, 2016). The second is modulation of synaptic transmission by retrograde inhibitory signaling (Alger, 2002; Regehr et al., 2009; Tao & Poo, 2001; Zilberter et al., 2005). This is a form of presynaptic inhibition, where postsynaptic cells release a neurotransmitter that binds to the receptors that are located on the presynaptic terminals. Retrograde signaling creates a feedback loop that dynamically regulates the amount of transmitter that is released from the presynaptic terminals. Here, we have hypothesized that such interactions occur in recurrent pathways from the excitatory nodes to the inhibitory interneuron and back from the interneuron to the excitatory nodes. In the excitatory-to-inhibitory pathway, retrograde signaling enables the inhibitory interneuron to compute the maximum instead of the sum of its inputs. Computation of the maximum arises from the limitation that the activity of the inhibitory interneuron cannot grow beyond the maximal input that it receives from the excitatory nodes. Furthermore, retrograde signaling in the inhibitory-to-excitatory pathway enables the excitatory nodes that receive maximal input to protect themselves from the common inhibition. In this way, the network can select all excitatory nodes with maximal input, irrespective of their quantity or arrangement in visual space.

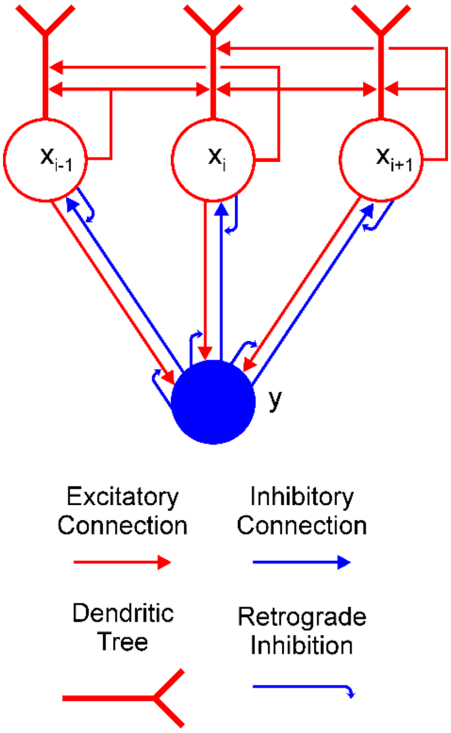


Figure 2. *Feature-based winner-take-all (F-WTA) circuit. Connections between excitatory (red circles) and inhibitory (blue disk) units are modulated by retrograde inhibition (curved blue arrows). Self-excitation and nearest-neighbor excitation are mediated by the dendrites of the excitatory units. The same motif is repeated for all excitatory nodes in the recurrent map.*

At first sight, it might appear strange to propose that an excitatory unit can inhibit its input by releasing a neurotransmitter that binds to the presynaptic terminal. However, several signaling molecules have been identified to support such interactions, including endogenous cannabinoids (Alger, 2002). Moreover, Zilberter (2000) found that glutamate is released from dendrites of pyramidal neurons in the rat neocortex and suppresses the inhibition that impinges on them. In addition, similar action has been found for GABA (Zilberter et al., 1999), which suggests that conventional neurotransmitters can engage in retrograde signaling.

To situate the proposed F-WTA circuit in a larger neural architecture that describes the cortical computations that underlie top-down attentional control, we have adopted the model that was proposed by Hamker (2004). He showed how attentional selection of a target arises from the recurrent interactions within a distributed network that consists of model cortical area V4, the inferotemporal cortex (IT), the posterior parietal cortex (PPC), and the frontal eye fields (FEF). Figure 3 illustrates part of these interactions that are involved in feature-based attentional guidance. Top-down signals that provide feature cues originate in the IT, which contains a spatially invariant representation of relevant visual features. The IT sends feature-specific feedback projections to the V4, where topographically organized feature maps for each feature value are located. For simplicity, we consider only maps for two colors (red and green) and two orientations (vertical and horizontal). We do not explicitly model IT and V4 dynamics. Rather, they serve here as a tentative explanation of how input to the F-WTA network arises within the ventral visual pathway. Also, we omitted the contribution of the FEF and its spatial reentry signals to the V4 activity.

We hypothesize that the feature-based WTA network resides in the PPC, where it receives summed input over all feature maps from the V4. Top-down guidance is implemented by a temporary increase in activity in one of the V4 feature maps. For example, when the decision is made to attend to the red color, the IT representation of red color sends feedback signals to the Red Map in the V4. Top-down signals to the feature map are modeled as a multiplicative gain of neural activity, which is consistent with neurophysiological findings (Martinez-Trujillo & Treue, 2004; Maunsell & Treue, 2006; Treue & Martinez-Trujillo, 1999).

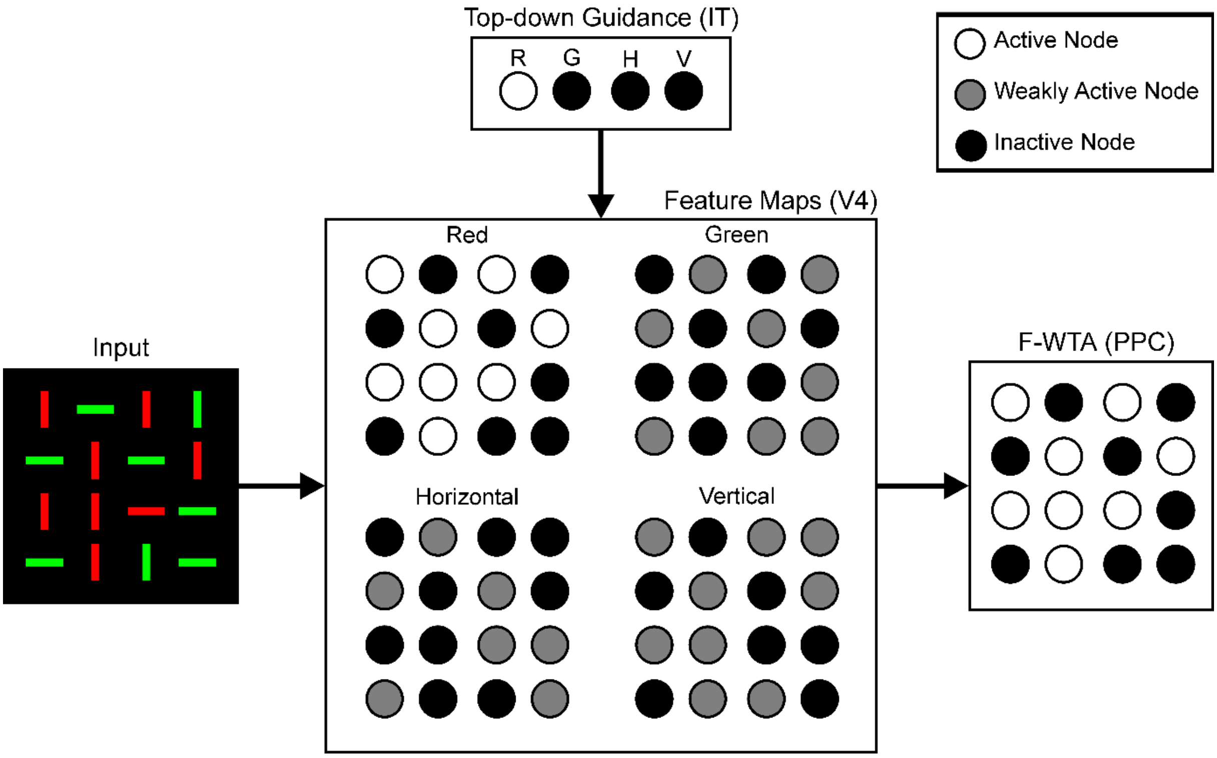


Figure 3. *Neural architecture for the top-down guidance of attention by feature cues, following Hamker (2004). Input is processed in retinotopically organized feature maps for colors and orientations. These maps also receive top-down signals, which provide feature guidance. In this example, input image is taken from Figure 1B and red color is cued by the top-down signals. Therefore, the activity of the nodes in the Red map is enhanced (white discs) relative to the activity in all other feature maps (gray discs) because latter receives only feedforward signals. Black discs represent inactive nodes. In the feature maps, this indicates the absence of a feature at a given locations. The F-WTA network sums output of all feature maps. Its activity represents all locations occupied by the cued feature. Parentheses contain reference to cortical areas thought to be involved in proposed computations. R – red; G – green; H – horizontal; V – vertical.*

The following neural network equations represent the quantitative description of the model. Each unit is defined by its instantaneous firing rate (Dayan & Abbott, 2000). The time evolution of the activity of excitatory node *x* at position *i* in the recurrent map is given by the following differential equation:

$$\tau_x \frac{dx_i}{dt} + x_i = \left[I_i(t) + \alpha f\left(x_i + x_{i+1} + x_{i-1}\right) - \beta_1 g\left(y - x_i - T_y\right) \right]^+ . \qquad (1)$$

The time evolution of the activity of inhibitory interneuron *y* is given by

$$\tau_y \frac{dy}{dt} + y = \left[\beta_2 \sum_i g\left(x_i - y - T_x\right) \right]^+ . \tag{2}$$

Parameters τ_x and τ_y are integration time constants for excitatory and inhibitory nodes, respectively. We assume that inequality $\tau_x > \tau_y$ holds, which accords with the observation in electrophysiological measurements that inhibitory cells exhibit faster dynamics than excitatory cells (McCormick et al., 1985). The second term on the left-hand side of Equations (1) and (2) describes the passive decay that drives the unit's activity to the resting state in the absence of external input. Firing rate activation function $[u]^+$ is a non-saturating rectification nonlinearity, which is defined by

$$[u]^+ = \max(u, 0). \tag{3}$$

Following Hamker (2004), we assume that feedforward input I_i at time t to the excitatory node x_i in the F-WTA network is given by the sum over activity in all V4 feature maps $I_i^{(m)}$,

$$I_i(t) = \sum_m I_i^{(m)} G^{(m)}(t). \tag{4}$$

In Equation (4), m denotes available feature maps with $m \in \{red, green\}$ in the simulation that is reported in section Simulation of the Formation of a Single Boolean Map and $m \in \{red, green, horizontal, vertical\}$ in the simulation that is reported in section Simulation of the Intersection and Union of Two Boolean Maps. Parameter G^m refers to the feature-specific, global multiplicative gain that all units $I_i^{(m)}$ within the same feature map m receive via top-down projections. As shown in Figure 2, these projections arrive from the feature representation in the IT. Multiplicative gating is generally consistent with previous models that describe the effect of feature-based attention on the responses of neurons in the early visual cortex (Boynton, 2005, 2009). Equation (4) ensures that the F-WTA network is not particularly sensitive to any feature value. Rather, it signals the behavioral relevance of locations in a spatial map. Here, the relevance can be set according to differences in the bottom-up input $I_i^{(m)}$ that arise from competitive interactions in the early visual cortex. Alternatively, relevance can be signaled by the top-down feature cues G^m that change the gain of all locations that are occupied by the same feature value.

Dendritic output $f(u)$ is described by the sigmoid response function

$$f(u) = \frac{S_d}{1 + e^{-\lambda(u - T_d)}} \tag{5}$$

where λ and T_d control the shape of the sigmoid function and S_d is its upper asymptotic value. We set λ to a high value to achieve a steep rise of the dendritic activity immediately after its input crosses the dendritic threshold, which is denoted as T_d. Such strong nonlinearity is justified by experimental data, which show all-or-none behavior in real dendrites (Wei et al., 2001; Polsky et al., 2004). In Equation (1), parameter α controls the strength of the impact that the dendritic compartment exerts on the soma.

Self-recurrent x_i and nearest-neighbor collaterals x_{i-1} and x_{i+1} arrive on the dendrite of the excitatory node, which is consistent with the anatomical observation that most recurrent excitatory connections are made on the dendrites of the excitatory cells (Spruston, 2008). Nodes at the edge of the network receive excitation only from a single available neighbor. That is, node x_1 receives excitation only from x_2, and x_N receives excitation only from x_{N-1}. Nearest-neighbor excitatory interactions enable feature cues to spread activity enhancement automatically to all connected locations that contain a given feature value. This is not essential for the simulation of Boolean maps, but we included it in our model because recurrent connections among nearby neurons are prominent feature of the synaptic organization of the cortex (Douglas & Martin, 2004). Also, we wanted to show that the proposed model is capable of simulating object-based attention (Roelfsema, 2006; Roelfsema & de Lange, 2016). Moreover, Wannig et al. (2011) found direct evidence for activity spreading among neurons that encode the same feature value in the primary visual cortex.

The output of the presynaptic interactions $g(u)$ is defined by the rectification nonlinearity of the form

$$g(u) = [u]^+ = \max(u, 0)\,. \tag{6}$$

In Equation (1), the term $-\,g(y - x_i - T_y)$ describes the output of the presynaptic terminal that delivers inhibition from interneuron y to excitatory node x_i (Figure 4A). However, we did not explicitly model the dynamics of retrograde signaling. We assumed that the release of the retrograde transmitter occurs simultaneously with the activation of the postsynaptic node and that it is proportional to its firing rate. Therefore, it is represented by the term – x_i.

Function $g(u)$ ensures that the presynaptic terminal will release the inhibitory transmitter only when the electrical signal from node y exceeds the inhibitory retrograde signal $–x_i$ and the threshold for presynaptic activation, which is denoted as T_y. In other words, node x_i will be inhibited only if $y > x_i + T_x$. If this is not the case, node x_i will effectively isolate itself from the inhibitory influence of node y. This is always the case for the winning node because $x(t) > y(t)$ for $t > 0$. Moreover, this result extends to all other nodes whose input magnitude is sufficiently close to the maximal input. The strength of the inhibition is determined by parameter β_1. In a similar vein, in Equation (2), the term $–g(x_i – y – T_x)$ describes the action of the retrograde signal that is released from inhibitory interneuron y on the presynaptic terminal that delivers excitation from node x_i (Figure 4B). Here, parameter T_x describes the threshold for the activation of the presynaptic terminal of the excitatory node and β_2 determines the strength of the excitation.

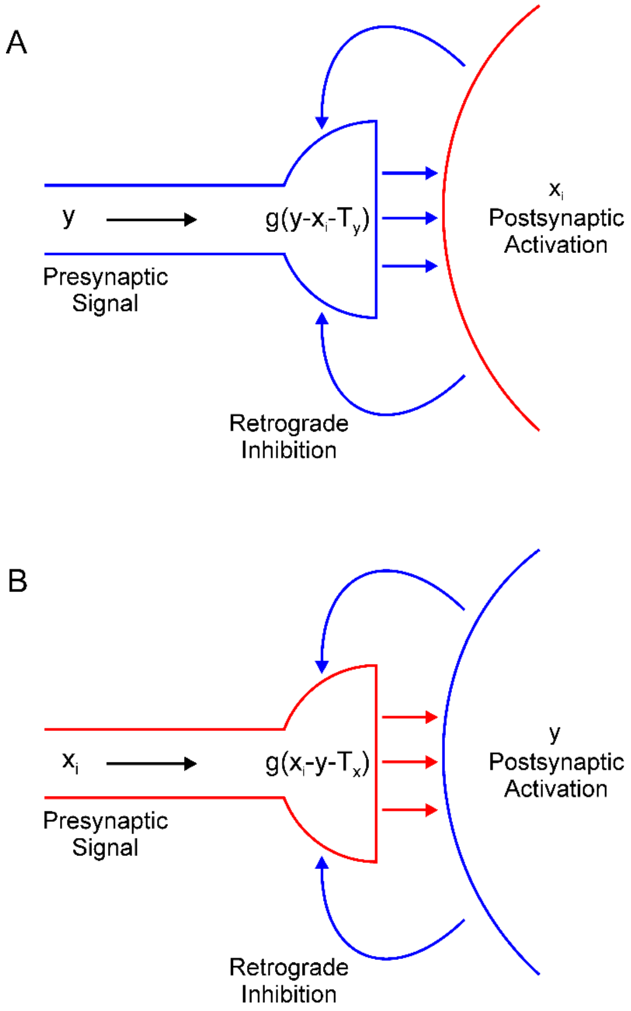


Figure 4. *Retrograde inhibitory signaling (blue curved arrows) from excitatory node x_i to the presynaptic terminal of inhibitory interneuron y (A) and from the inhibitory interneuron to the presynaptic terminal of the excitatory node (B). Both terminals compute half-wave rectification g(u) of their input. Terminals release respective inhibitory (A) or excitatory (B) neurotransmitter (straight horizontal arrows) only when they receive net positive input.*

We have proposed a model of a one-dimensional network, although it attempts to simulate phenomena that occur in 2-D, as illustrated by Figure 1. We have chosen to work with the 1-D version of the network simply because we want to focus on the analysis of its temporal dynamics and its ability to combine information over time. Without loss of generality, the computer simulations that are reported in section Computer Simulations should be considered as a cross-section of a 2-D network.

For simplicity, the thresholds that control the activation of the excitatory and inhibitory nodes are all set to zero and are omitted from the model description. Parameters were set as follows: τ_x = 5; τ_y = 2; α = 1; β_1 = 1; β_2 = 10; S_d = 1; λ = 100; T_d = 0.1; T_x = 0.1; and T_y = 0.1.

Parameters were chosen in a way to simultaneously achieve intersection and union. Systematic variations on the parameters α, β_1 and β_2 showed that intersection is observed when $1 \leq (\alpha, \beta_1) \leq 5$. In contrast, union is observed when $0.8 \leq (\alpha, \beta_1) \leq 1$. Parameter β_2 can be set to any value above the default without changing the results.

3. MODEL EXTENSIONS

The network that is defined by Equations (1) and (2) is chosen in a way that achieves the desired behavior with the minimal number of computational elements. This simplicity heuristic is important for understanding model properties without adding extra neuroscientific complexity (Ashby & Hélie, 2011). However, at the same time, this approach sacrifices anatomical and biophysical plausibility of the proposed model. In this section, we present several extensions and generalizations of the basic model that bring it closer to satisfying the neurobiological constraints.

3.1. Inhibitory Pool

The model has just one inhibitory interneuron for computational convenience, which is not realistic. It is known that excitatory neurons outnumber inhibitory neurons by a factor of four in the cortex (Braitenberg & Schüz, 1991). However, it is possible to design an F-WTA network with a pool of inhibitory interneurons and the appropriate ratio between excitatory and inhibitory nodes that achieves the same behavior as the original model. An extended F-WTA network is presented in Figure 5A. Here, each inhibitory interneuron receives input from a subset of the excitatory nodes. We depicted each excitatory subset as a vertical arrangement of four nodes that do not overlap in their projections to the inhibitory pool. Therefore, each excitatory node projects to just one inhibitory node. Naturally, this does not need to be the case. It is possible that each excitatory node projects to more than one node without compromising the network output. Importantly, all inhibitory interneurons are mutually connected. In addition, each inhibitory interneuron projects its output to all excitatory nodes (denoted by thick blue arrow). As in the original model, we assume that all inhibitory and excitatory nodes are endowed with the capability of retrograde signaling on their synaptic contacts.

Within the pool of inhibitory nodes, retrograde signaling enables computation of the MAX function, as in the original model. To see this, consider the inhibitory node that receives maximal input. Due to the retrograde signaling, it will reach a steady state that corresponds to the computation of the MAX function over input from its excitatory subset. Moreover, it will not receive inhibition from the other members of the pool. All other inhibitory nodes, which receive less excitatory support, will be silenced because their retrograde signaling is not sufficiently strong to prevent lateral inhibition from the winning node. However, if there are multiple inhibitory nodes with the same level of activity, they will remain active together. Finally, the winning nodes send inhibition to all excitatory subsets. Since excitatory nodes also engage in retrograde signaling, the nodes that receive maximal input will block inhibition and remain active. Therefore, the network output will look much like the original model because the MAX computation on the inhibitory nodes makes irrelevant the number of them that are active simultaneously.

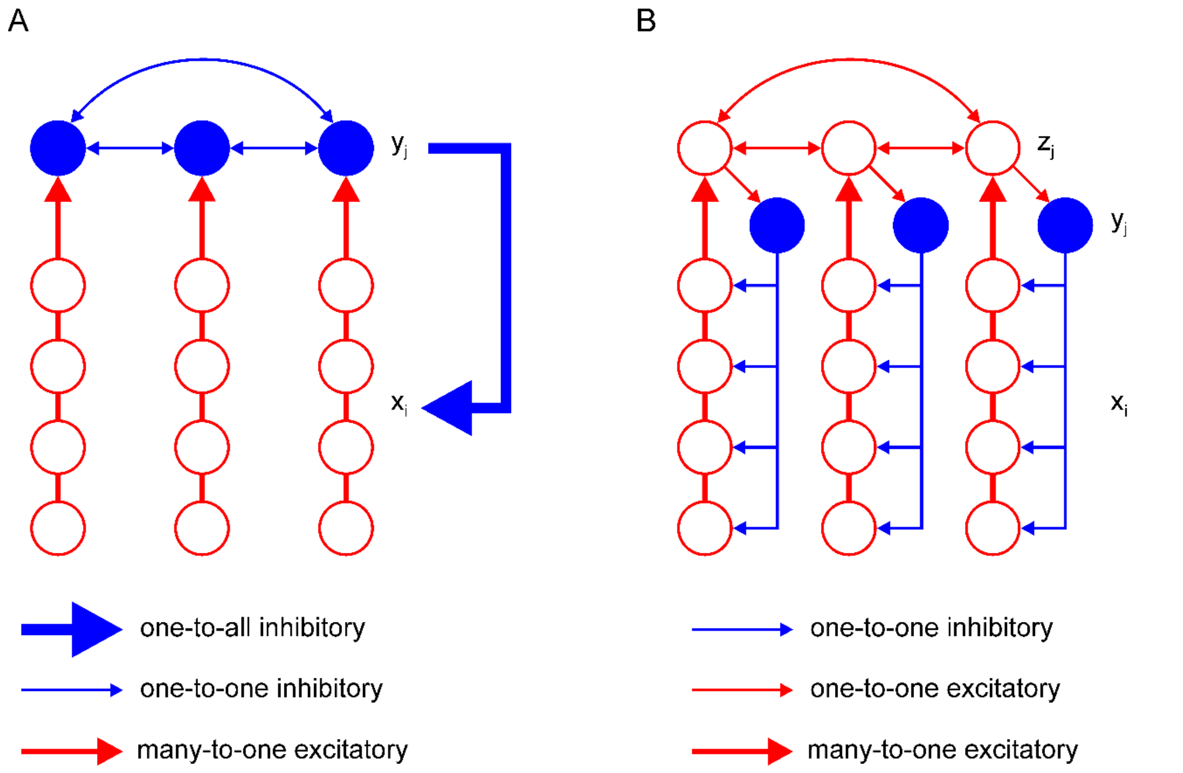


Figure 5. *Two variations of the F-WTA circuit design that are computationally equivalent to the basic circuit that is shown in Figure 2. (A) Circuit with a set of inhibitory nodes, which are denoted as y_j. Each y_j receives input from a subset of excitatory nodes. Inhibitory nodes compete with one another and the winning node encodes the maximum of its input. It delivers inhibition to all excitatory nodes in the same way as single inhibitory node y in the basic circuit. (B) Circuit with an additional set of excitatory nodes z_j with long-range horizontal projections. These nodes propagate the locally computed maximum level of activity to all parts of the network. Therefore, the whole set of z_j converges to a global maximum. Furthermore, they contact inhibitory nodes y_j that deliver inhibition to a subset of excitatory nodes x_i.*

3.2. Localized Inhibition

An important shortcoming of the previous model is that it assumes that inhibitory projections extend across the whole network of excitatory units. This is clearly not the case in real neural networks, where the spatial spread of inhibition is limited. To account for this property, we have constructed a more elaborate version of the basic model, which is shown in Figure 5B. It contains a new pool z_j of excitatory nodes with long-range projections. The z_j nodes receive input from the subset of the x_i nodes. Additionally, each z_j node sends its projection to at least one y_i node from the pool of inhibitory nodes. The number of z nodes must equal the number of inhibitory nodes y_j so that they can be indexed by the same subscript j.

Again, we assume that the z_j nodes are equipped with the ability of retrograde signaling on their synapses. Therefore, they also compute the MAX function over all their inputs, including feedforward input from the corresponding subset of x_i nodes and recurrent input from other z_j nodes. In this design, the maximum level of activity that is sensed by the x_i nodes in one part of the network is easily propagated via z_j nodes to all other parts of the network. Furthermore, z_j nodes transfer this activity to inhibitory nodes. Therefore, each inhibitory node will eventually receive the maximal level of activity and apply it to the subset of x_i nodes to which it is connected. In this design, it is not necessary for inhibitory nodes to interact with one another. The excitatory nodes x_i that receive maximal input will block inhibition by their retrograde signaling and remain active in the same manner as described in the previous section. In this way, the proposed circuit achieves the same result as the original model.

3.3. Output Functions

The model employs threshold-linear output functions for the soma and the logistic sigmoid function for dendrites. This is inconsistent with the observation that somatic output also saturates and is also often modeled by the sigmoid function. However, in normal circumstances, neurons operate in a linear mode that is far from their saturation level (Rutishauser & Douglas, 2009). To provide a more systematic approach to the output functions that are used in the model, we introduce a piecewise-linear approximation to the sigmoid function $s_q(u)$ of the form

$$s_q(u) = \begin{cases} 0 & if \quad u \le 0 \\ u & if \quad 0 < u < S_q \\ S_q & if \quad u \ge S_q \end{cases} \tag{7}$$

where S_q denotes the upper saturation point, which can be set differently for different computational units $q \in \{c, d, p\}$, which correspond to the somatic, dendritic, and presynaptic terminal outputs, respectively. With the output function $s_q(u)$ applied to all computational elements of a single node, the model equations, namely, Equations (1) and (2), can be restated as

$$\tau_x \frac{dx_i}{dt} + x_i = s_c \left[I_i(t) + \alpha s_d \left(x_i + x_{i+1} + x_{i-1} - T_d \right) - \beta_1 s_p \left(y - x_i - T_y \right) \right] \tag{8}$$

and

$$\tau_y \frac{dy}{dt} + y = s_c \left[\beta_2 \sum_i s_p \left(x_i - y - T_x \right) \right]. \tag{9}$$

An important constraint of the model that is defined by Equations (8) and (9) is that saturation point for the dendritic output S_d should be chosen to be smaller than S_c, which is the saturation point of the somatic output. In this way, feedforward input I_i can be combined with the dendritic output without causing saturation at the output of the node. In contrast, if dendrites are allowed to saturate at the same activity level as the node, the dendritic output will overshadow the feedforward input. Consequently, the network will lose its sensitivity to the input changes. This is undesirable with respect to the requirements that are imposed by the sequential formation of the multiple Boolean maps. Therefore, the choice between the linear or the sigmoid output function for the node is not important if the dendritic output is restricted to a smaller interval relative to the output of the node itself.

4. LINEAR STABILITY ANALYSIS

4.1. Fixed Points

Fixed point is found iteratively starting from the set of nodes receiving maximal input, x_M. We assume that the winning nodes and inhibitory interneuron are activated above their thresholds, so we set $[u]^+ = u$. Next, we observe that the winning nodes do not receive inhibition from the interneuron y since $x_M(t) > y(t)$ for $t > 0$. This holds because the activity of the inhibitory node is bounded above by $x_M + T_x > y$ where T_x is a positive constant. Then, retrograde signaling ensures that $g(y - x_M - T_y) = 0$ for all times t. Consequently, nodes receiving maximal input are driven solely by excitatory terms. Since the recurrent excitation is bounded above by its asymptotic value S_d, dendritic output function $f(u)$ in Equation (1) is replaced with S_d. This yields the following approximation to the steady state of the winning nodes:

$$x_M \approx I_M + \alpha S_d \ . \tag{10}$$

After the x_M, inhibitory interneuron y also reaches its steady state because its activity is driven primarily by the input from x_M. As the activity of y grows, terms $g(x_i - y - T_x)$ in Equation (2) vanish for all nodes that do not receive maximal input x_i where $i \notin M$. In contrast, the presynaptic terminals of x_M are above the threshold for their activation just before y reaches equilibrium, that is, $g(x_M - y - T_x) > 0$. Therefore, the output function of the presynaptic terminal $g(u)$ can be replaced by u. Then, Equation (2) is solved as

$$y = \frac{\beta_2 k\left(x_M - T_x\right)}{\beta_2 k + 1} \ , \tag{11}$$

where k is the number of x_M. When β_2 is chosen to be sufficiently large, and/or there are many nodes with maximal input x_M, then

$$y \to x_M - T_x \ . \tag{12}$$

Continuity of the function defined by Equation (2) implies that y cannot grow above $x_M - T_x$, that is, $y(t) > x_M\,(t) - T_x$ cannot hold at any time t unless $y(t_0) = x_M\,(t_0) - T_x$ at some earlier time $t_0 < t$. However, equality $y(t_0) = x_M\,(t_0) - T_x$ implies that $dy/dt = 0$ at time t_0 because $g(x_M\,(t_0) - y(t_0) - T_x) = 0$. In other words, node y loses all its excitatory drive when it reaches $x_M - T_x$. This is true irrespective of the number k of x_M. Thus, node y computes the maximum over its input.

The x_M nodes, together with the inhibitory node, create a quenching threshold (QT) for the network, which is defined by

$$QT = y - T_y = x_M - T_x - T_y \ . \tag{13}$$

Grossberg (1973) introduced the concept of the quenching threshold to describe the property of contrast enhancement in recurrent competitive networks. Nodes whose activity is above QT are enhanced and stored in the memory state, while all nodes whose activity is below QT are suppressed and removed from the memory representation. In the same manner, the remaining excitatory nodes converge to one of two states, depending on whether they exceed QT or not:

$$x_{i \notin M} \approx \begin{cases} I_i + \alpha S_d & if \quad x_i \geq QT \\ 0 & if \quad x_i < QT \end{cases} . \tag{14}$$

QT and its relationship with the activity of the winning and non-winning nodes and inhibitory interneuron is illustrated in Figure 6. According to Equations (10), (11), and (14), the fixed-point linearly combines input and recurrent excitation. As maximal input increases or decreases, the fixed point will move up or down and track these changes. Moreover, the input may cease, and the winning nodes will settle into the activity level that is provided by the recurrent excitation alone, which is expressed as *αS_d*. In other words, the network remembers who the last winner was. The same is true in the case where the winner is determined by transient cues that are applied sequentially on a sustained input. This is a protocol that is used in the computer simulations that are reported in section Computer Simulations.

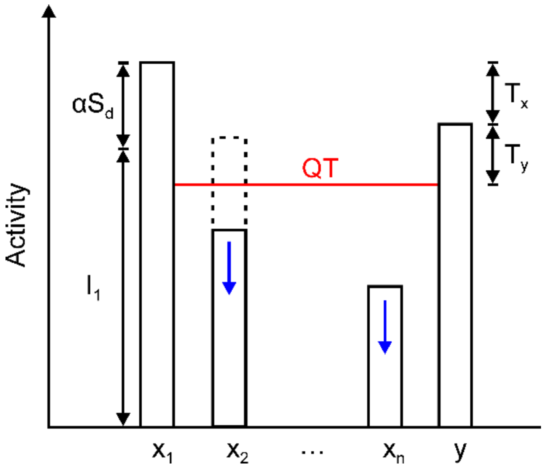


Figure 6. *Relationship among the steady state of the winning node x_1, inhibitory node y, and all other excitatory nodes in the network, x_2 ... x_n. The activity of the winning node is given by the sum of its feedforward input I_1 and the output of its dendrite mediating self- and nearest neighbor excitation, which is expressed as αS_d. Inhibitory node y approximately converges to $x_1 - T_x$. It sets the quenching threshold (QT) that separates excitatory nodes into two sets. Nodes x_2 ... x_n are spared from inhibition if their activity is above the QT (dashed line); otherwise, they are silenced to zero (solid line). QT equals $y - T_y$ (or $x_1 - T_x - T_y$) because the activity of the inhibitory node must exceed the threshold on its presynaptic terminals that contact the excitatory nodes.*

4.2. Linearization Near Fixed Points

To simplify the stability analysis, we consider an F-WTA network with two excitatory nodes and one inhibitory node: [x_1, x_2, y]. This system has three fixed points: x_1 is the only winner, x_2 is the only winner, and both excitatory nodes are winners. To which fixed point the network will converge depends on the relationship between inputs I_1 and I_2.

Local stability of the fixed point is estimated from the eigenvalues of the Jacobian matrix, which is the matrix of partial derivatives of the system of equations. If the real parts of all eigenvalues of the Jacobian are negative, the fixed point will be asymptotically stable (Rutishauser & Douglas, 2009). However, before we can compute the Jacobian matrix, we note that a linear-threshold function is continuous, but not differentiable. To sidestep this problem, we follow the approach that was described by Rutishauser et al. (2011) of inserting dummy terms that correspond to the derivate. That is, we need three separate dummy terms: c_i and p_{xi}, which correspond to the somatic and presynaptic output functions of excitatory node i, and a set of p_{yi} dummy terms that describe the presynaptic output function of inhibitory node y. The dummy terms are defined as

$$c_i = p_{xi} = p_{yi} = \frac{d}{du}\left[u_i(t)\right]^+ = \begin{cases} 0 & if \quad u_i(t) \le 0 \\ 1 & if \quad u_i(t) > 0 \end{cases}. \tag{15}$$

Based on the above definition of the dummy terms, we have constructed the Jacobian matrix of the system that consists of Equations (1) and (2):

$$J = \begin{bmatrix} \tau_x^{-1}\left(c_1\left(\alpha D_1 f + \beta_1 p_{y1}\right) - 1\right) & \tau_x^{-1} c_1 \alpha D_2 f & \tau_x^{-1} c_1 \beta_1 p_{y1} \\ \tau_x^{-1} c_2 \alpha D_1 f & \tau_x^{-1}\left(c_2\left(\alpha D_2 f + \beta_1 p_{y2}\right) - 1\right) & \tau_x^{-1} c_2 \beta_1 p_{y2} \\ \tau_y^{-1} \beta_2 p_{x1} & \tau_y^{-1} \beta_2 p_{x2} & -\tau_y^{-1}\left(\beta_2\left(p_{x1} + p_{x2}\right) - 1\right) \end{bmatrix} \tag{16}$$

where D_1f and D_2f denote the partial derivatives of the sigmoid function with respect to x_1 and x_2. Now, we examine the Jacobian matrix at the three fixed points that are mentioned above. If x_1 is the only winner, then $c_1 = 1$. However, $D_{x1}f \approx 0$ because the recurrent excitation of the winning node approaches its asymptotic value, which is S_d. In addition, $p_{y1} = 0$ because the winning node blocks inhibition from node y, as discussed above. Node x_2 is inhibited below its somatic threshold, that is, $c_2 = 0$. Presynaptic signaling by inhibitory node y blocks excitation from x_1 and x_2 is inactive, so $p_{x1} = p_{x2} = 0$. Consequently, the Jacobian matrix at the fixed point reduces to a diagonal matrix of the form

$$J_{W1} = J_{W2} = J_{W12} = \begin{bmatrix} -\tau_x^{-1} & 0 & 0 \\ 0 & -\tau_x^{-1} & 0 \\ 0 & 0 & -\tau_y^{-1} \end{bmatrix}. \tag{17}$$

All eigenvalues of the J_{W1} are negative, and the fixed point is asymptotically stable. In the case when x_2 is the sole winner, the same arguments are applied to set the dummy terms, thereby leading to the same diagonal matrix J_{W2} as shown in (17). Moreover, if both excitatory nodes are winners, then $c_1 = c_2 = 1$, $D_{x1}f = D_{x2}f \approx 0$ and $p_{x1} = p_{x2} = 0$. Again, the Jacobian matrix J_{W12} is diagonal. Thus, all three fixed points are asymptotically stable.

The same analysis can be generalized to a network of arbitrary size and arbitrarily many fixed points. Retrograde signaling and dendritic saturation will ensure that the Jacobian matrix of any size will be diagonal and that the network dynamics will be independent of the network parameters, namely, α, β_1, and β_2. Local stability analysis suggests that the system behaves much like a feedforward network that is driven by the input. However, an important difference is that the F-WTA network has memory states like the recurrent network (Rutishauser & Douglas, 2009; Usher & Cohen, 1999).

5. COMPUTER SIMULATIONS

We performed a set of computer simulations to illustrate the model behavior. We employed a vector of 200 excitatory units and one inhibitory unit. Differential Equations (1) and (2) were solved numerically using MATLAB’s *ode15s* solver. The simulations were run for 250 time steps. In subsequent figures, we followed the convention that activity of the node at position *i* as a function of time is depicted by a shade of gray, with white representing the maximal value and black representing zero.

5.1. Simulation of the Formation of a Single Boolean Map

First, we demonstrate how a Boolean map arises in the F-WTA network in response to the presentation of the color cue, as illustrated by Figure 1A. In Figure 7A, we recreate a similar stimulus condition in the 1-D map. The input consists of red and green items of equal sizes, which are intermixed in space on a black background. Input magnitude *I* was set to 1 in both maps and to 0.2 in the empty space around items to represent spontaneous activity in the absence of visual stimulation. Initially, the top-down or attentional gain is set to $G^m = 1$ in both feature maps $m \in \{red, green\}$. At $t = 50$, the red color is attended, which is reflected in the input to the network by increasing the gain for all nodes in the Red map ($G^{red} = 2$) and simultaneously reducing the gain in the Green map by the same factor ($G^{green} = 1/G^{red} = 1/2$). Top-down gain is also applied to the empty space between items, which is consistent with the finding that feature-based attention spreads across the whole visual field (Saenz et al., 2002, 2003; Serences & Boynton, 2007). The duration of the top-down cue is 50 simulated time steps. For simplicity, top-down signals are suddenly switched on and off without exponential decay. At $t = 150$, the green color is cued in the same way.

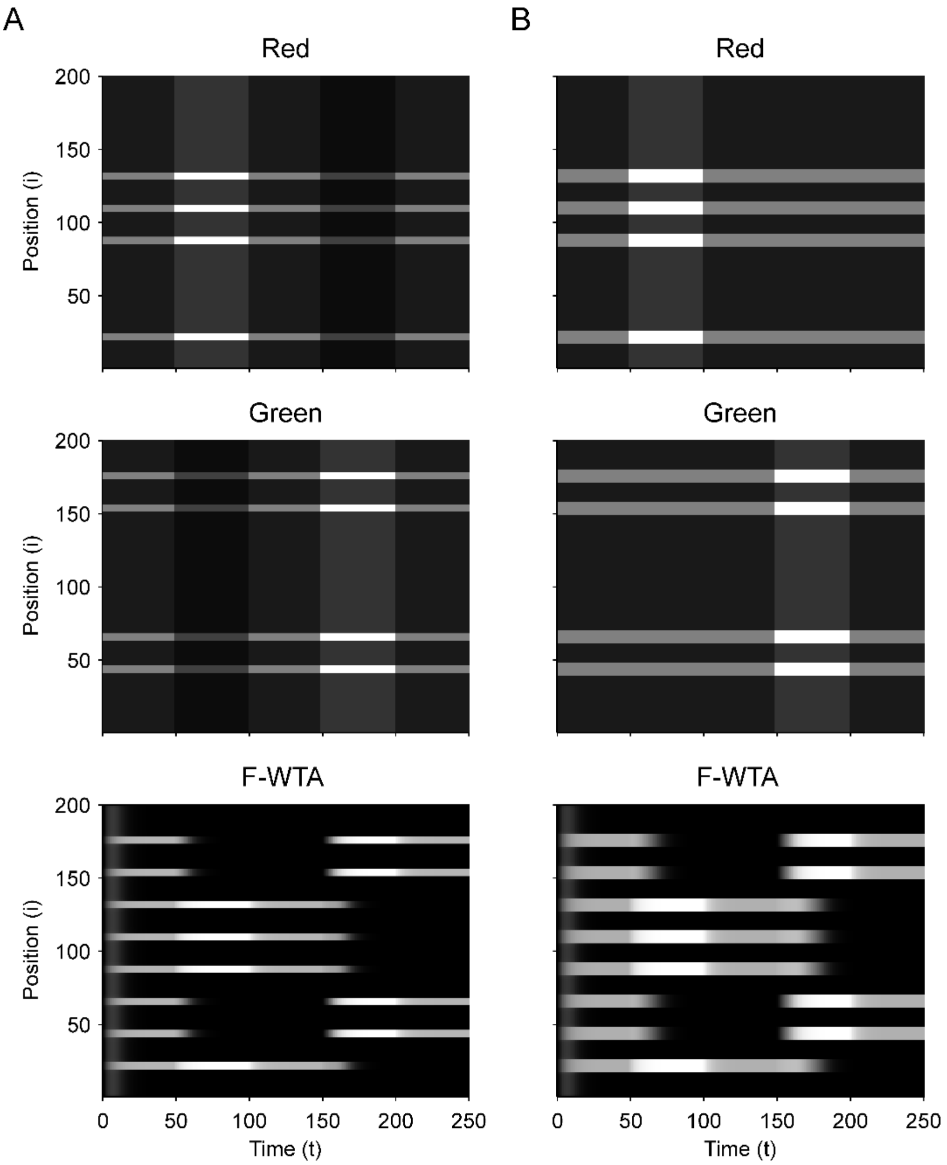


Figure 7. *(A) Simulation of the Boolean map formation in the F-WTA network in response to the sequential presentation of two color cues (red appeared between 50th and 100th and green appeared between 150th and 200th simulated time unit). (B) The same simulation with larger items and without gain modulation applied on the unattended feature map.*

At the beginning of the simulation, before the top-down signals are applied, the F-WTA network simply selects all presented items together, irrespective of their color. Next, when the red color is cued by applying top-down signals to the corresponding feature map, the network responds to the new input by selectively increasing and sustaining the activity of nodes that encode locations of red items in the input and suppressing locations that encode green items. That is, the network creates a Boolean map by highlighting the spatial pattern that is associated with the red color. Furthermore, due to a self-excitation, the network maintains locations of the cued feature value in working memory after the top-down signals cease to influence the feature map. When the observer decides to switch attention to another feature value, the network can select the locations of the new feature value and suppress the locations that are associated with the previously cued value without requiring an external reset. Namely, the network is sensitive to input changes even though it also exhibits activity persistence.

Importantly, the activity level at selected locations is invariant with respect to the number of active nodes. At the beginning of the simulation, the number of active nodes was four times larger than after the cue was delivered. However, the active nodes remained at the same activity level as they were at the beginning of the simulation. This is a consequence of retrograde inhibitory signaling in recurrent pathways. It prevents unbounded growth of inhibition due to the dynamic regulation of its strength. To illustrate this point further, we run another simulation with items that are almost double in size (Figure 7B). Even though the total size of the cued items is increased, the activity of the cued nodes converges to the same level as before. In this simulation, we also checked that the network successfully operates even if we remove gain reduction from the non-attended feature map.

Next, we determined the minimal feature gain that must be applied on the input to produce the desired behavior. When the gain modulation is applied simultaneously on attended feature map G^A and on unattended feature map G^{NA} (where $G^{NA} = 1/G^A$), we found that $G^A \geq 1.7$ is sufficient for creating a Boolean map and switching to another one. In contrast, when the gain modulation is not applied on the unattended feature map, as shown in Figure 7B, the feature gain in the attended map should be set to $G^A \geq 2$ to achieve the same behavior.

Figure 8 illustrates that the F-WTA network can support space- and object-based attention alongside feature-based attention. When the spatial cue is applied to a single location in one of the feature maps, the network responds by selecting only this location. Neighboring nodes are not selected even though they are reciprocally connected to the cued node. The reason is that they receive weaker input relative to the cued node. Furthermore, recurrent excitation that arrives from the cued node is bound by the dendritic nonlinearity. Thus, it is not sufficiently strong to keep them active. Interestingly, when the spatial cue is removed, the network activity starts to propagate from the cued node towards the boundary of the whole item. In this case, the network selects not just the cued location, but all locations that are connected to it. Therefore, the F-WTA network exhibits object-based selection, which is consistent with neurophysiological studies that show spreading of enhanced activity along the shape of the object (Roelfsema, 2006). This property arises because the removal of the cue equalizes the input magnitude along the object, which allows activity enhancement to propagate via local lateral connections. In addition, this simulation shows that spatial attention can be easily oriented toward a new location in a single jump without the need for attentional pointers that move attention across the map (Hahnloser et al., 1999).

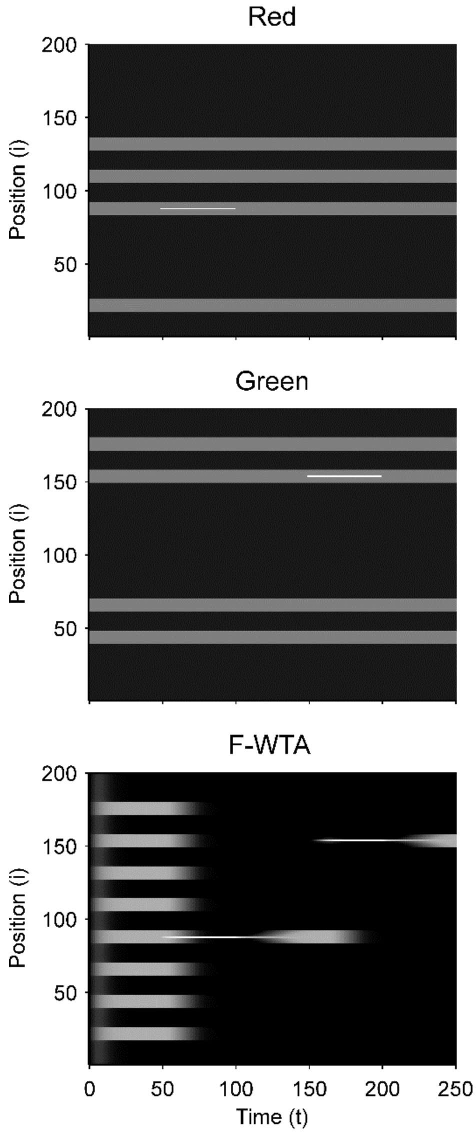


Figure 8. *Simulation of space- and object-based attention in the F-WTA network.*

5.2. Simulation of the Intersection and Union of Two Boolean Maps

Figure 9 illustrates that the model can sequentially combine two Boolean maps when the network is cued by top-down signals from two separate feature dimensions. In this simulation, we have employed a visual input that consists of red and green horizontal and red and green vertical bars, like those that are illustrated in Figure 1B. First, the F-WTA network is cued to select red bars, irrespective of their orientation. In the second step, it is cued to select horizontal bars, irrespective of their color. However, green vertical bars are already suppressed and the top-down signal that is supplied to them is not sufficient to override the inhibition that arises from red vertical bars. The net result is the selection of a subset of red horizontal bars. In other words, the network activity converges to an intersection between a set of red bars and a set of horizontal bars, thereby resulting in the selection of red horizontal bars.

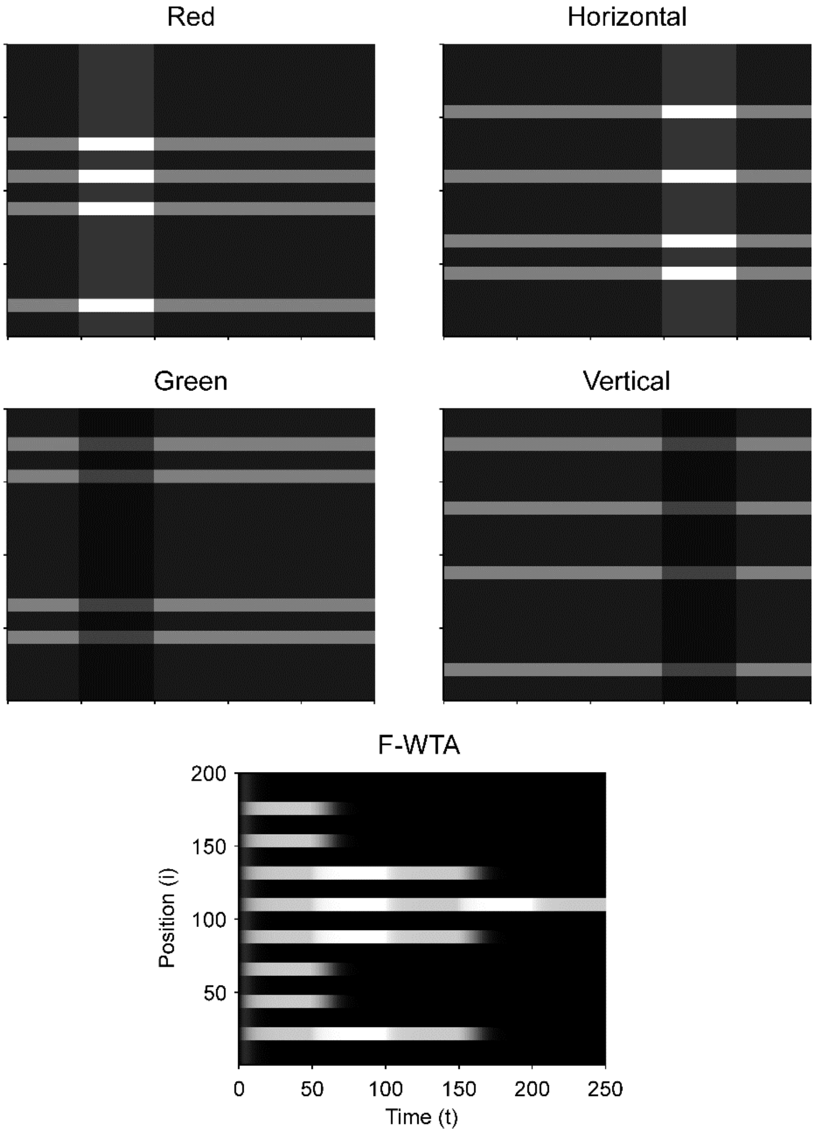


Figure 9. *Simulation of the intersection of red and horizontal items.*

Next, we examined how the network achieves the union of two Boolean maps (Figure 10). Here, we assumed that the input consists of two non-overlapping components: colored squares that activate color maps but do not activate orientation maps, and achromatic horizontal and vertical bars that activate orientation maps but do not activate color maps, as shown in Figure 1C. Red-colored items occupy locations between 1 and 100 and oriented bars occupy locations between 101 and 200. This closely resembles the stimulus that is used by Huang and Pashler (2007) to demonstrate the union of color and texture. Taken together, the data show that the union of two Boolean maps is possible only when two top-down cues overlap in time or when the second cue closely follows the withdrawal of the first cue. In Figure 10, the cue for the red map is applied in the interval [50, 100] and the cue for the horizontal map is applied in the interval [110, 160]. In this case, the F-WTA network converges to the union of red and horizontal items. However, when top-down cues do not overlap, as shown in Figure 11, the second cue overrides the network activity that remains from the first cue. We suggest that this property partly explains why the union is difficult to achieve, as observed by Huang and Pashler (2007).

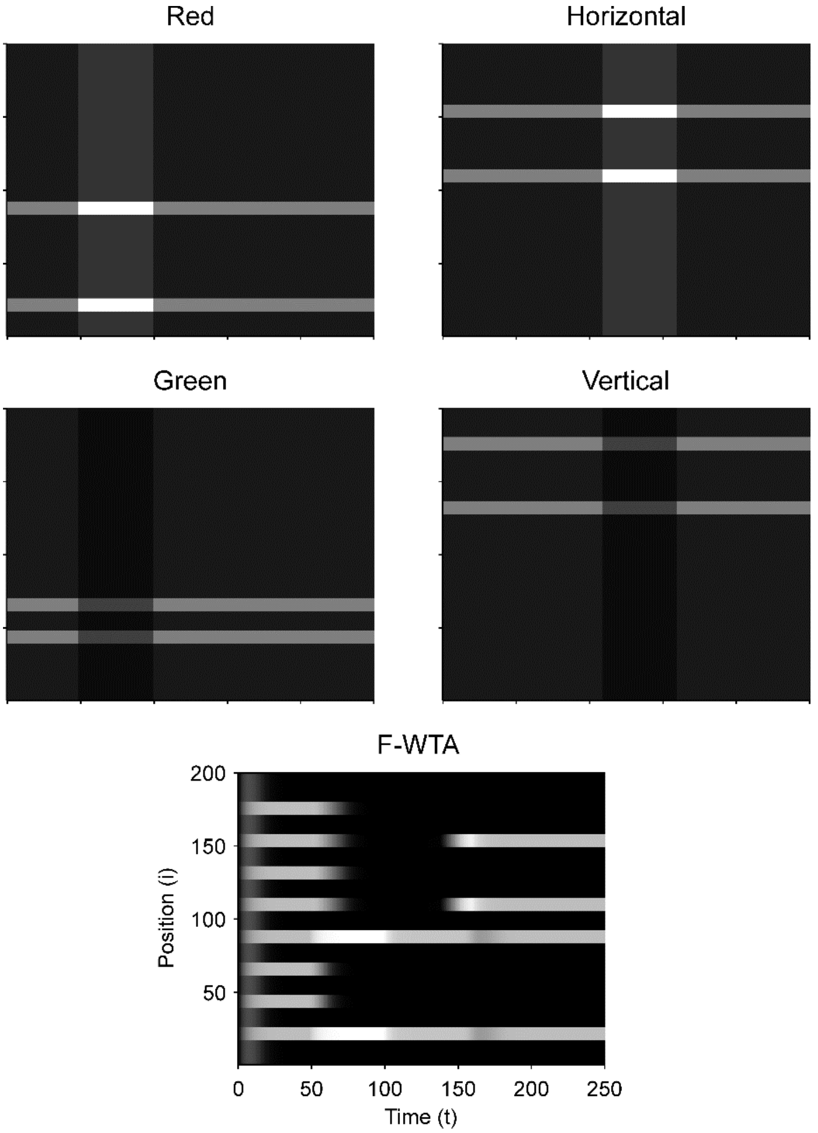


Figure 10. *Simulation of the union of red and horizontal items.*

In addition, we examine the boundary conditions on the choice of the feature gain parameter. We parametrically vary the feature gain in steps of 0.1 starting from $G = 2$ and moving below and above to determine when the ability to form the intersection or union breaks down. When the gain modulation is applied simultaneously on attended (G^A) and unattended (G^{NA}) feature maps, we find that G^A should be chosen from the interval [1.5, 2.1] to achieve the intersection between two maps. When $G^A < 1.5$, the network fails to segregate cued from non-cued locations in the first step. In contrast, when $G^A > 2.1$, the network successfully segregates cued from non-cued locations in the first step. However, the gain is too high, so all horizontal items are selected together in the second step. That is, the representation of red horizontal items is merged with the representation of green horizontal items. When $G^{NA} = 1$ throughout the simulation, G^A should be chosen from the interval [1.8, 2.0] to achieve intersection.

With respect to the union of two maps, the feature gain G^A should be chosen from the interval [1.4, 2.0] when $G^{NA} = 1/G^A$ and from the interval [1.6, 2.0] when $G^{NA} = 1$. When G^A is chosen below the suggested intervals, feature gain is too weak, and the second cue will not be able to raise the activity level of the nodes that represent horizontal items above the quenching threshold. Therefore, the network ends up with the Boolean map of red items that is formed in the first step. When G^A is chosen above the suggested interval, the network switches between the representation of the red items in the first step to the representation of the horizontal items in the second step. In this case, the feature cue is too high, and the activity of the nodes that represent horizontal items simply overrides the activity of the nodes that represent the red items. These constraints are derived from the situation in which the two top-down cues overlap in time. As shown above, temporal lag of the second cue relative to the first cue also destroys the ability of the network to form the union of two Boolean maps.

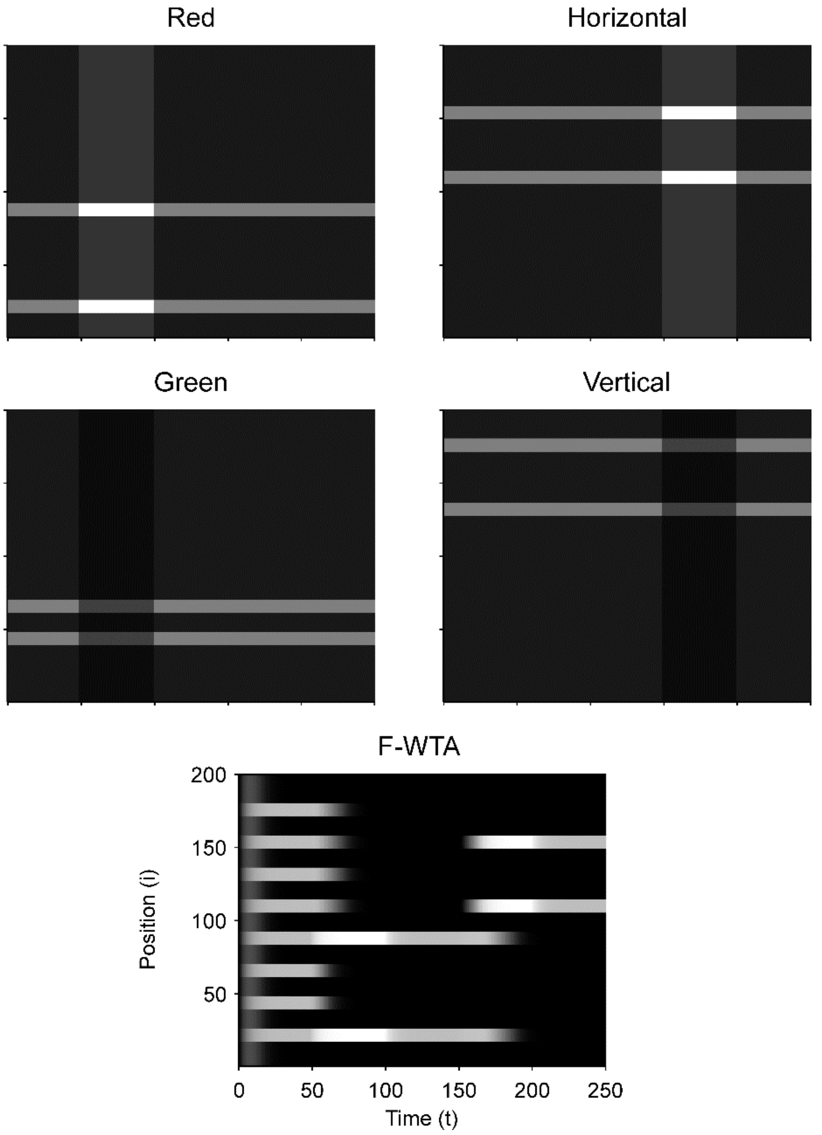


Figure 11. *Breakdown of union of red and horizontal items when the delivery of top-down cues for red and horizontal items is separated by a large temporal gap.*

5.3. Simulation of Bottom-up Spatial Selection

Finally, we have shown that when there is no top-down guidance, the network selects the most-salient locations based on the bottom-up salience that is computed within feature maps (Figure 12). We did not explicitly model competition among maps, but it is reasonable to assume that in a scene with many multi-featured objects, their input magnitudes (i.e., saliencies) will be different. Therefore, we arbitrarily assigned different input magnitudes to different items. As shown in Figure 12A, the F-WTA network selects the most salient object if the difference in input magnitude between the two most active nodes is sufficiently large. However, when this difference is small, as shown in Figure 12B, the F-WTA model chooses two most salient items together. Furthermore, in both examples, the network activity retains the input amplitude of the winning item (or items), thereby illustrating the ability to compute the function maximum (Yu et al., 2002).

The precision of saliency detection depends on the threshold for the activation of synaptic receptors on the inhibitory interneuron. In all reported simulations, it was set to T_y = 0.1. If smaller values were chosen, the network would improve in terms of precision and be able to separate the two objects that are presented in Figure 12B. However, this comes at the price of losing the ability to form a union of two Boolean maps. Therefore, there is a trade-off between the precision of saliency detection and the ability to form Boolean maps.

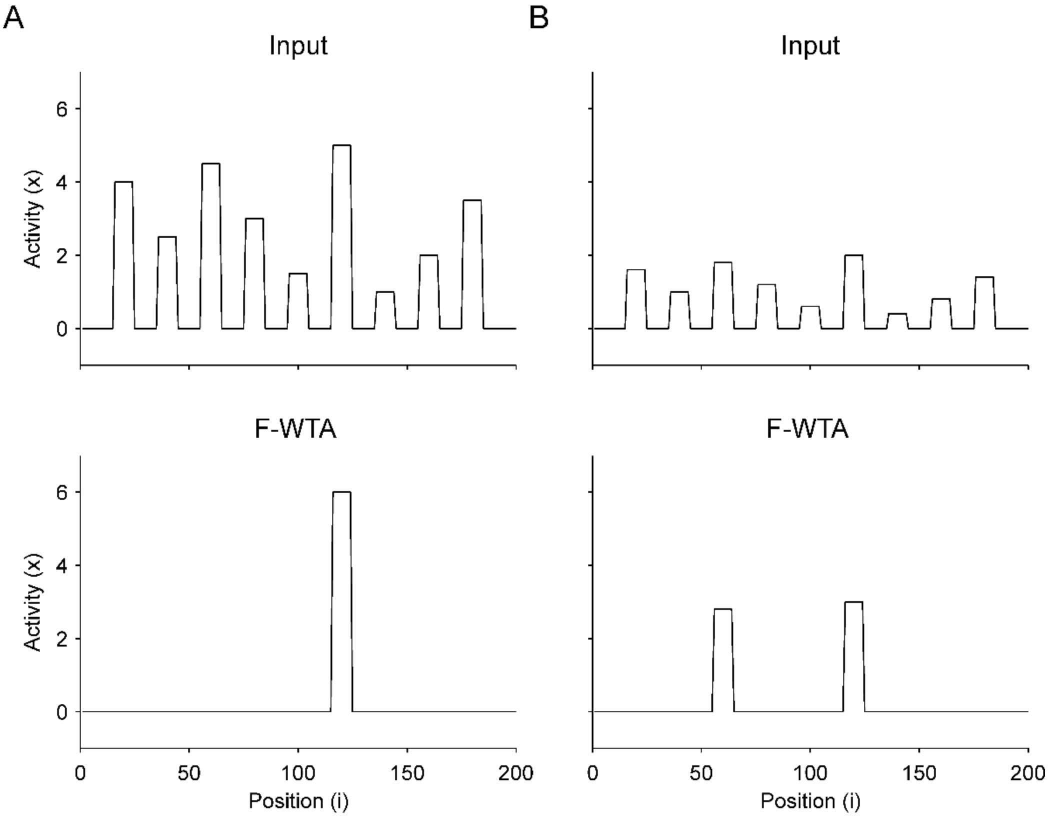


Figure 12. *Selection of the most salient item in the absence of top-down guidance. (A) When the most salient item is sufficiently distinctive from other items, the F-WTA network selects it. (B) When the saliency of all items is relatively low, the F-WTA network may select more than one item because it has a limit on the precision by which it separates inputs of different magnitudes.*

An important aspect of stimulus-driven attentional control is attentional capture by peripheral cues. Behavioral studies have shown that the abrupt onset of a new object in a visual scene can automatically capture attention even if it is irrelevant for the current goal (Theeuwes, 2010). Figure 13 illustrates the sensitivity of the F-WTA network to abrupt visual onset. To simulate this effect, we have made the additional assumption that the network receives input not only from a sustained channel that is comprised of feature maps in V4 but also from a transient channel that responds vigorously only to changes in input (Kulikowski & Tolhurst, 1973; Legge, 1978). Thus, when the abrupt onset is accompanied by a strong transient signal that exceeds the activity level of the currently attended item, the F-WTA network temporarily switch activity towards the location of the onset (Figure 13A). Here, the input at the locations that are occupied by the winning item in the center of the map was set to $I_W = 2$. Input to all other items was set to $I_i = 1$. Finally, the transient input that appears on the sides of the map was set to $I_T = 4$. It is sufficient to set $I_T \geq I_W + 0.8$ to achieve sensitivity to abrupt onsets. Moreover, the same relation holds even if we choose a larger value for I_W.

Next, when abrupt onset produces only weak transient signals ($I_T = 2$) that do not satisfy the inequality that is stated above ($I_W = 2$), the activity in the F-WTA network resists abrupt onset and stays on the previously attended item (Figure 13B). This observation is consistent with behavioral findings that abrupt onset can be ignored (Theeuwes, 2010), perhaps by attenuating the response of the transient channel. Another possibility is that the top-down gain for the attended location can be increased so that it exceeds the activity of the transient channel. In this case, intense focus on the current object prevents attentional capture, which is consistent with the psychological concept of the attentional window (Belopolsky & Theeuwes, 2010).

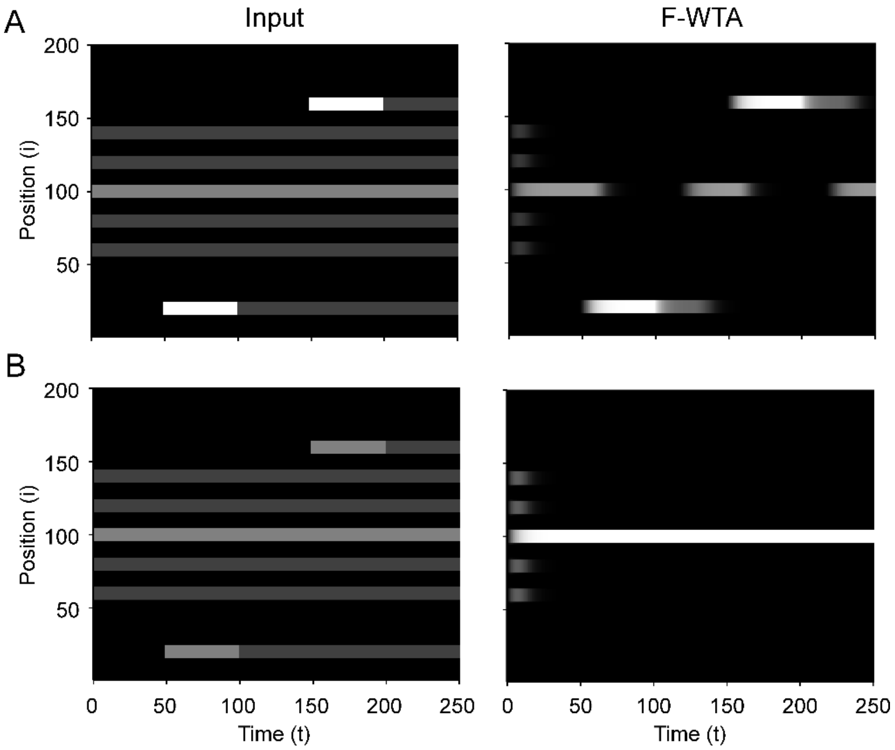


Figure 13. *Sensitivity to abrupt visual onsets. (A) When the transient signal that is produced by the abrupt onset of a new object is sufficiently strong, it temporarily draws attention to itself. (B) When the transient signal is weak, attention resists abrupt onset and stays on the item that was selected at the beginning of the simulation.*

6. DISCUSSION

We have proposed a new model of the WTA network that can simultaneously select multiple spatial locations based on a shared feature value. We named the model the feature-based WTA (F-WTA) network because the unit of selection is not a point in space or object, but rather an abstract feature value that is set by the top-down signals. We have demonstrated how the F-WTA network implements the central proposal of the Boolean theory of visual attention that there exists a spatial map that divides the visual space into two mutually exclusive sets. One set represents all locations that are occupied by the chosen feature value. The other set contains all other locations, which are not of interest. The Boolean map controls spatial selection and access to the consciousness (Huang & Pashler, 2007). Moreover, we have shown that the network successfully integrates information across space and time to form the intersection or union of two maps that are defined by different feature cues. Previous models of the WTA network are not capable of such integration because they require that the current winner be externally inhibited to allow attentional focus to move from one location to another (Itti & Koch, 2000, 2001; Kaski & Kohonen, 1994). Another possibility to move activity across locations in the network is to introduce dynamic thresholds that simulate habituation or fatigue in individual neurons. In this case, current winner loses its competitive advantage due to the raise of its threshold. This allows non-winners to gain access to working memory (Horn & Usher, 1990). However, both approaches are not suitable for forming the intersection or union of a set of previous winners and a set of later winners.

Another important property of the F-WTA network that sets it apart from previous models of WTA behavior is the ability to select and store arbitrarily many locations in the memory. This is achieved by inhibitory retrograde signaling, which effectively isolates winning nodes from mutual inhibition. First, the amount of inhibition in the network is significantly reduced because the inhibitory interneuron computes the maximum instead of the sum of the recurrent input that it receives from the excitatory nodes. Second, the winning excitatory nodes release their retrograde signals and block inhibition from the interneuron. Consequently, arbitrarily many winners can participate in representing the selected locations without degrading their activation. In other words, there is no capacity limit on the number of objects that can be simultaneously selected. This is consistent with recent behavioral findings that suggest that our ability to select multiple objects is not fixed. Rather, spatial attention should be considered a fundamentally continuous resource without a strict capacity limit (Alvarez & Franconeri, 2007; Davis et al., 2000; Davis et al., 2001; Liverence & Franconeri, 2015; Scimeca & Franconeri, 2015).

In addition, the network is sensitive to the sudden appearance of a new object in the scene, which suggests that it can also be guided by bottom-up feature cues (Theeuwes, 2013). We hypothesize that the network receives strong input from the transient channel. Such input overrides the network’s current memory state, thereby making it sensitive to abrupt onsets. Moreover, the transient channel can be activated by any type of change in the spatiotemporal energy of the input, and not just by the sudden appearance (or disappearance) of objects. For example, it will be activated by a sudden change in the direction of motion (Farid, 2002). When the network simultaneously receives transient input from different locations, they all will be selected together. In this way, the network achieves temporal grouping of synchronous transient input. That is, the network can discover spatial structures that are defined purely by temporal cues (Lee & Blake, 1999; Rideaux et al., 2016).

6.1. Biophysical Considerations

As noted above, the model of the F-WTA network rests upon three key computational elements: the dendrite as an independent computational unit, retrograde signaling on synaptic contacts, and computing the maximum over inputs. Here, we review supporting neuroscientific evidence that suggests that all three biophysical mechanisms are plausible candidates for computation in real neural networks.

There is a growing body of evidence that the excitatory pyramidal cell should not be viewed as a single electrical compartment. Rather, it consists of multiple independent synaptic integration zones arranged in a two-layer hierarchy (Branco & Häusser, 2010; Häusser & Mel, 2003; London & Häusser, 2005; Mel, 2016). Using a detailed biophysical model of the pyramidal neuron, Poirazi et al. (2003) showed that its output is well approximated by a two-layer neural network. In the first layer of the network, dendrites independently integrate their synaptic input and produce sigmoidal output. In the second layer, the dendritic output is summed at the soma to produce the neuron’s firing rate. Importantly, the somatic and dendritic output functions need not be the same (Jadi et al., 2014). For example, Behabadi and Mel (2014) showed that the soma of the model neuron generates nearly linear output, while the dendritic output is sigmoid. In our model, the dendrite conveys recurrent excitation to the node. Due to the dendritic nonlinearity, there is no risk of unbounded activity growth in the node. Furthermore, the dendritic output is summed with the external input at the soma of the node. By using a linear output function at the soma, we have ensured that the F-WTA network remains sensitive to input fluctuations.

Synaptic transmission can be dynamically regulated in an activity-dependent manner, as shown by the existence of depolarization-induced suppression of inhibition (DSI) (Pitler & Alger, 1992) and depolarization-induced suppression of excitation (DSE) (Kreitzer & Regehr, 2001). DSI (DSE) refers to the reduction in inhibitory (excitatory) post-synaptic potentials following depolarization of the postsynaptic cell. These processes have been observed in various brain regions, including the cerebellum, hippocampus, and neocortex. A retrograde messenger that is released from postsynaptic cell due to its depolarization mediates DSI and DSE. After release, the retrograde messenger binds to the receptors at the presynaptic axon terminals and suppresses the release of the transmitter. Based on these properties, Regehr et al. (2009) suggested that a possible physiological function of DSI and DSE is to provide negative feedback that reduces the impact of the synaptic input on the ongoing neural activity.

The model behavior rests upon the assumption that the inhibitory interneuron computes the maximum instead of the sum of its inputs. There is some direct physiological evidence that real cortical neurons indeed compute the MAX function. For example, Sato (1989) examined responses of neurons in the primate inferior temporal cortex to the presentation of one or two bars in their receptive field. He concluded that the responses to two bars that were presented simultaneously were well described by the maximum of the responses to each separately. In a similar vein, Gawne and Martin (2002) recorded the activity of neurons in primate V4 and found that their firing rate in response to the combination of stimuli is best described by the maximum function over the firing rates that are evoked by each stimulus alone. Furthermore, Lampl et al. (2004) directly measured membrane potentials in the complex cells of the cat primary visual cortex and found evidence for the MAX-like behavior in response to the pair of optimal bars.

Indirectly, the importance of the MAX-like operation in cortical information processing can be appreciated by considering the many computational models of visual functions that have employed it in simulating rich and complex datasets. For example, Riesenhuber and Poggio (1999) employed hierarchical computation of the MAX function in a model of invariant object recognition. Spratling (2010, 2011) used it in simulating a large range of classical and non-classical receptive field properties of V1 neurons. Moreover, Tsui et al. (2010) used MAX-like input integration to explain diverse properties of MT neurons and Hamker (2004) used it in his model of top-down guidance of spatial attention. Furthermore, Kouh and Poggio (2008) developed a canonical cortical circuit that is capable of many nonlinear operations, including computation of the MAX function. Here, we have shown that a single inhibitory node that is endowed with retrograde signaling can compute the maximum.

Based on the proposed model, we have derived two testable predictions. The cortical network that is involved in spatial selection will contain inhibitory interneurons that can compute the MAX function. Moreover, both the excitatory and inhibitory neurons in this network will be endowed with the anatomical structures that support retrograde signaling (presynaptic receptors and postsynaptic transmitter release sites).

6.2. Comparison with Other WTA Network Models

Several models of biophysical mechanisms have been proposed for implementing WTA behavior in a neural network, including linear-threshold units (Hahnloser, 1998; Rutishauser &

Douglas, 2009), non-linear shunting units (Fukai & Tanaka, 1997; Grossberg, 1973), and oscillatory units (Borisyuk & Kazanovich, 2004; Wang, 1999).

A simple model of a competitive network that is based on linear-threshold units has been extensively studied. Stability analysis revealed that this network requires fine-tuning of the connectivity to achieve stable dynamics that can perform cognitively relevant computations, such as choice behavior (Hahnloser, 1998; Hahnloser et al., 2003; Rutishauser et al., 2015). Recently, Binas et al. (2014) showed that a biophysically plausible learning mechanism could tune the network connections in a way that keeps the network dynamics in the stable regime. Here, we have shown how dendritic and synaptic nonlinearities ensure that the network dynamics near fixed points depends only on the time constants of the nodes and not on the parameters that control recurrent excitation and lateral inhibition. Therefore, a precise balance between excitation and inhibition is not necessary for achieving a stable memory state. Moreover, the network is sensitive to the input and can iteratively combine the current memory state with new input to form the intersection or union of them.

An important problem for WTA networks that are based on the linear-threshold or sigmoid output functions is that they lack a mechanism for controlling inhibition between the winning nodes. Therefore, they have limited capacity to represent multiple winners. Usher and Cohen (1999) showed that their activation decreases up to the point of complete inactivation as the number of winning nodes increases. This is due to the increased amount of mutual inhibition. The problem cannot be solved simply by reducing the strength of the lateral inhibition because it is not known in advance how many locations will be cued. On the other hand, feature-based spatial selection requires that the network be able to adjust automatically the amount of inhibition to accommodate the selection of a very small or very large number of winners.

Grossberg (1973) proposed a recurrent competitive map model that was based on shunting non-linear interaction between the synaptic input and the membrane potential. The output of the model depends on the exact form of the signal function that is used to convert membrane potential into the firing rate. When the signal function is chosen to grow faster than linear, the network exhibits WTA behavior. By contrast, when the signal function is sigmoid, the network can select multiple winners if they have similar activity levels. The most important property of this model is the existence of the quenching threshold. All nodes whose activity is above QT are enhanced and all nodes whose activity is below QT are suppressed. This behavior is similar to the operation of the F-WTA network that was proposed here. However, an important difference is that in the shunting model, QT is fixed and dependent on the parameters of the network. In contrast, the feature-based WTA network exhibits dynamic QT that depends on the input to the network and not on its parameters. In this way, the F-WTA network rescales its sensitivity to the input fluctuations.

More recently, a version of the recurrent competitive map was applied in modeling object-based attention (Fazl et al., 2009). It was shown that sustained network activity in the model PPC encompasses the whole object as an attentional shroud around it. Such spatial representation of a single object supports view-invariant object recognition within a larger neural architecture, namely, ARTSCAN. In an extension of the model, Foley et al. (2012) proposed two separate competitive networks that account for distinct properties of object- and space-based attention. A network with strong inhibition is limited to the selection of a single object. The other network utilizes weaker inhibition to support multifocal spatial selection. To increase the capacity of this network to represent multiple objects, Foley et al. (2012) suggested that the amount of lateral inhibition could be controlled externally. As the number of objects that should be selected together increases, the lateral inhibition should become weaker to counteract the effect of the larger number of nodes that participate in the competition. In contrast, the F-WTA network does not require such external adjustments of the strength of the lateral inhibition to accommodate the selection of arbitrarily many objects of arbitrary size. Moreover, in the F-WTA network, object-based and multifocal spatial attention coexist within the same circuit. Whether the network exhibits object-based spatial selection depends on the type of cue that is presented to the network and not on its parameters.

Wang (1999) proposed a model of object-based attention that relies on the phase synchronization and desynchronization among oscillatory units. At each location of the recurrent map, there is a pair of excitatory and inhibitory units with distinct temporal dynamics that creates a relaxation oscillator. Excitatory units are also mutually connected with their nearest neighbors and with a global inhibitor. The network is initialized with random phase differences between oscillators at different network locations. The activity of the global inhibitor further enforces phase separation among excitatory units. However, local excitatory interactions among nearest neighbors oppose global inhibition and result in phase synchronization that spreads among nodes that encode the same object. The net result of these interactions is temporal segmentation and selection of one active object representation at a time in a multi-object input image. Importantly, the network can switch its activity from one object representation to another. However, this transition is generated internally by the oscillator dynamics. It is not possible to drive the object selection by external cues such as top-down gain control or bottom-up cues such as abrupt onsets. Moreover, it is not possible to enforce simultaneous selection of more than one object by a joint feature value because the global inhibitor will desynchronize all nodes that encode non-connected items. Therefore, it is not clear how synchronous oscillations could support feature-based attentional selection. Taken together, it is still an open issue whether they are relevant for perception and cognition (Ray & Maunsell, 2015).

6.3. Limitations

The proposed model of spatial selection successfully simulates the formation of the Boolean map and its elaboration by the set operations of intersection and union but does not fully implement all aspects of the theory that was proposed by Huang and Pashler (2007). Precisely, it does not explain why attention is limited to only one feature value per dimension or how the observer sequentially chooses one feature value after another or combines feature dimensions into intersections or unions of Boolean maps. It is likely that this severe limitation arises from some form of the WTA network. However, this constraint requires a more elaborate model of the interactions among the spatially invariant representation of the feature values in the IT cortex and the interactions between the IT and the prefrontal cortex, where decisions and plans are made.

In all simulations that are reported here, we kept items segregated in space. This was not the case in the stimuli that were used by Huang and Pashler (2007). They employed a matrix of colored squares that were connected to one another. This is because activity spreading can occur among adjacent nodes even if they encode different feature values. Activity spreading is observed after top-down signals stop favoring one feature value over the other. In this case, all feature maps contribute equally to the input of the F-WTA network and the network is no longer able to discriminate between selected and unselected feature values. One way to solve this issue is to assume that the top-down signals are constantly present during the whole trial. In this way, the activity magnitude on the cued locations is kept above that on the non-cued locations. Therefore, non-cued locations are treated as background noise and suppressed, despite their proximity to the cued locations. Another possibility is to impose boundary signals that act upon recurrent collaterals of the nodes in the F-WTA network in a way that is similar to how activity spreading is stopped in the network models of brightness perception (Grossberg & Todorović, 1988), visual segmentation (Domijan, 2004), and figure-ground organization (Domijan & Šetić, 2008).

Finally, input to the network does not follow the distance-dependent activity profile that is usually observed in the visual cortex. However, this is not a critical issue for the model's performance because the precision of selection depends on the thresholds for presynaptic terminal activation, namely, $T_{x,}$ and T_y. If they are set to very small values, the network will tend to select the centers of the objects when the input pattern is convolved with a Gaussian filter. In contrast, if they are set to larger values, the network will be able to select extended parts of the objects and possibly even the whole objects. In the same way, the model achieves resistance to the input noise. As thresholds are set to larger values, the network can tolerate a larger amount of noise. However, this comes at a cost of less-precise selection, as demonstrated by the simulation that is shown in Figure 12.

7. CONCLUSIONS

We have demonstrated how the feature-based WTA network achieves spatial selection of all locations that are occupied by the same feature value without suffering from capacity limitations. The network responds to the top-down cue by storing in memory spatial pattern that corresponds to the cued feature value, while non-cued feature values are suppressed. In this way, we have shown how the Boolean map is formed. In addition, we have shown that it is possible to create more complex spatial representations that involve the intersection or the union of two or more Boolean maps. In this way, the F-WTA network goes beyond the capabilities of previous models of the competitive neural network, which cannot integrate information across space and time. Our work suggests that dendritic nonlinearity and retrograde signaling are biophysically plausible mechanisms that are essential for model success.

FUNDING

This research was supported by the Croatian Science Foundation Research Grant HRZZ-IP-11-2013-4139 and the University of Rijeka Grant 13.04.1.3.11.

REFERENCES

Alger, B. E. (2002). Retrograde signaling in the regulation of synaptic transmission: focus on endocannabinoids. *Progress in Neurobiology, 68*(4), 247–286. https://doi.org/10.1016/S0301-0082(02)00080-1

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13), 14.1–10. https://doi.org/10.1167/7.13.14

Ashby, F. G., & Hélie, S. (2011). A tutorial on computational cognitive neuroscience: Modeling the neurodynamics of cognition. *Journal of Mathematical Psychology, 55*, 273–289. https://doi.org/10.1016/j.jmp.2011.04.003

Behabadi, B. F., & Mel, B. W. (2014). Mechanisms underlying subunit independence in pyramidal neuron dendrites. *Proceedings of the National Academy of Sciences*, *111*(1), 498–503. https://doi.org/10.1073/pnas.1217645111

Belopolsky, A. V., & Theeuwes, J. (2010). No capture outside the attentional window. *Vision Research, 50*(23), 2543–2550. https://doi.org/10.1016/j.visres.2010.08.023

Binas, J., Rutishauser, U., Indiveri, G., & Pfeiffer, M. (2014). Learning and stabilization of winner-take-all dynamics through interacting excitatory and inhibitory plasticity. *Frontiers in Computational Neuroscience, 8*(68). https://doi.org/10.3389/fncom.2014.00068

Borisyuk, R. M., & Kazanovich, Y. B. (2004). Oscillatory model of attention-guided object selection and novelty detection. *Neural Networks, 17*(7), 899–915. https://doi.org/10.1016/j.neunet.2004.03.005

Boynton, G. M. (2005). Attention and visual perception. *Current Opinion in Neurobiology, 15*(4), 465–469. https://doi.org/10.1016/j.conb.2005.06.009

Boynton, G. M. (2009). A framework for describing the effects of attention on visual responses. *Vision Research, 49*(10), 1129–1143. https://doi.org/10.1016/j.visres.2008.11.001

Braitenberg, V., & Schüz, A., (1991). *Anatomy of the cortex. Statistics and geometry* (Vol. 18). Springer-Verlag.

Branco, T., & Häusser, M. (2010). The single dendritic branch as a fundamental functional unit in the nervous system. *Current Opinion in Neurobiology*, *20*(4), 494–502. https://doi.org/10.1016/j.conb.2010.07.009

Davis, G., Driver, J., Pavani, F., & Shepherd, A. (2000). Reappraising the apparent costs of attending to two separate visual objects. *Vision Research*, *40*(10–12), 1323–1332. https://doi.org/10.1016/S0042-6989(99)00189-3

Davis, G., Welch, V. L., Holmes, A., & Shepherd, A. (2001). Can attention select only a fixed number of objects at a time? *Perception*, *30*(10), 1227–1248. https://doi.org/10.1068/p3133

Dayan, P., & Abbott, L. F. (2000). *Theoretical neuroscience: Computational and mathematical modeling of neural systems.* MIT Press.

Domijan, D. (2004). Recurrent network with large representational capacity. *Neural Computation*, *16*(9), 1917–1942. https://doi.org/10.1162/0899766041336422

Domijan, D., & Šetić, M. (2008). A feedback model of figure-ground assignment. *Journal of Vision*, *8*(7):10, 1–27. https://doi.org/10.1167/8.7.10

Douglas, R., & Martin, K. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, *27*, 419–451. https://doi.org/10.1146/annurev.neuro.27.070202.144152

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517. https://doi.org/10.1037/0096-3445.113.4.501

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 32–39. https://doi.org/10.1037/0096-1523.10.1.32

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/bf03211502

Farid, H. (2002). Temporal synchrony in perceptual grouping: A critique. *Trends in Cognitive Sciences*, *6*(7), 284–288. https://doi.org/10.1016/S1364-6613(02)01927-7

Fazl, A., Grossberg, S., & Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology*, *58*(1), 1–48. https://doi.org/10.1016/j.cogpsych.2008.05.001

Foley, N. C., Grossberg, S., & Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: Object cueing, useful-field-of-view, and crowding. *Cognitive Psychology*, *65*(1), 77–117. https://doi.org/10.1016/j.cogpsych.2012.02.001

Fukai, T., & Tanaka, S. (1997). A simple neural network exhibiting selective activation of neuronal ensembles: From winner-take-all to winners-share-all. *Neural Computation, 9*(1), 77–97. https://doi.org/10.1162/neco.1997.9.1.77

Gawne, T. J., & Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *Journal of Neurophysiology, 88*(3), 1128–1135. https://doi.org/10.1152/jn.00151.200

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52*, 217–257. https://doi.org/10.1002/sapm1973523213

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87*(1), 1–51. https://doi.org/10.1037/0033-295X.87.1.1

Grossberg, S., & Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception & Psychophysics, 43*(3), 241–277. https://doi.org/10.3758/bf03207869

Haarmann, H., & Usher, M. (2001). Maintenance of semantic information in capacity limited item short-term memory. *Psychonomic Bulletin & Review, 8*(3), 568–578. https://doi.org/10.3758/bf03196193

Hahnloser, R. L. (1998). On the piecewise analysis of networks of linear threshold neurons. *Neural Networks, 11*(4), 691–697. https://doi.org/10.1016/S0893-6080(98)00012-4

Hahnloser, R., Douglas, R. J., Mahowald, M., & Hepp, K. (1999). Feedback interactions between neuronal pointers and maps for attentional processing. *Nature Neuroscience, 2*, 746–752. https://doi.org/10.1038/11219

Hahnloser, R. H., Seung, H. S., & Slotine, J.-J. (2003). Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Computation, 15*(3), 621–638. https://doi.org/10.1162/089976603321192103

Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research*, *44*, 501–521. https://doi.org/10.1016/j.visres.2003.09.033

Häusser, M., & Mel, B. W. (2003). Dendrites: Bug or feature? *Current Opinion in Neurobiology*, *13*(3), 372–383. https://doi.org/10.1016/S0959-4388(03)00075-8

Horn, D., & Usher, M. (1990). Excitatory–inhibitory networks with dynamical thresholds. *International Journal of Neural Systems*, *1*(3), 249–257. https://doi.org/10.1142/S0129065790000151

Huang, L. (2015). Grouping by similarity is mediated by feature selection: Evidence from the failure of cue combination. *Psychonomic Bulletin & Review, 22*(5), 1364–1369. https://doi.org/ 10.3758/s13423-015-0801-z

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114*(3), 599–631. https://doi.org/10.1037/0033-295X.114.3.599

Huang, L., & Pashler, H. (2012). Distinguishing different strategies of across-dimension attentional selection. *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 453–464. https://doi.org/10.1037/a0026365

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*, 194–203. https://doi.org/10.1038/35058500

Jadi, M., Behabadi, B. F., Poleg-Polsky, A., Schiller J., & Mel, B. W. (2014). An augmented two-layer model captures nonlinear analog spatial integration effects in pyramidal neuron dendrites. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers, 102*(5), 782–798. https://doi.org/10.1109/jproc.2014.2312671

Kaptein, N. A., Theeuwes, J., & van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 1053–1069. https://doi.org/10.1037/0096-1523.21.5.1053

Kaski, S., & Kohonen, T. (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Networks*, *7*, 973–984. https://doi.org/10.1016/S0893-6080(05)80154-6

Kouh, M., & Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations. *Neural Computation*, *20*(6), 1427–1451. https://doi.org/10.1162/neco.2008.02-07-466

Kreitzer A. C., & Regehr, W. G. (2001). Retrograde inhibition of presynaptic calcium influx by endogenous cannabinoids at excitatory synapses onto Purkinje cells. *Neuron, 29*(3), 717–727. https://doi.org/10.1016/S0896-6273(01)00246-X

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology, 232*(1), 149–162. https://doi.org/ 10.1113/jphysiol.1973.sp010261

Lampl, I., Ferster, D., Poggio, T., & Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *Journal of Neurophysiology*, *92*(5), 2704–2713. https://doi.org/10.1152/jn.00060.2004

Lee, S. H., & Blake, R. (1999). Visual form created solely from temporal structure. *Science, 284*(5417), 1165–1168. https://doi.org/10.1126/science.284.5417.1165

Legge, G. E. (1978). Sustained and transient mechanisms in human vision: Temporal and spatial properties. *Vision Research*, *18*(1), 69–81. https://doi.org/10.1016/0042-6989(78)90079-2

Liverence, B. M., & Franconeri, S. L. (2015). Resource limitations in visual cognition. In R. Scott and S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–13). John Wiley and Sons.

London, M., & Häusser, M. (2005). Dendritic computation. *Annual Review of Neuroscience, 28*(1), 503–532. https://doi.org/10.1146/annurev.neuro.28.061604.135703

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, *14*(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

McCormick, D. A., Connors, B. W., Lighthall, J. W., & Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, *54*(4), 782–806. https://doi.org/10.1152/jn.1985.54.4.782

Mel, B. W. (2016). Towards a simplified model of an active dendritic tree. In G. Stuart, N. Spruston, & M. Häusser (Eds.), *Dendrites. Third edition* (pp. 465–486). Oxford University Press.

Nobre, A. C., & Kastner, S. (2014). *The Oxford handbook of attention*. Oxford University Press.

O'Grady, R. B., & Müller, H. J. (2000). Object-based selection operates on a grouped array of locations. *Perception and Psychophysics*, *62*(8), 1655–1667. https://doi.org/10.3758/bf03212163

Pitler, T. A., & Alger, B. E. (1992). Postsynaptic spike firing reduces synaptic GABAA responses in hippocampal pyramidal cells. *Journal of Neuroscience, 12*(10), 4122–4132.

Poirazi, P., Brannon, T. M., & Mel, B. W. (2003). Pyramidal neuron as two-layer neural network. *Neuron, 37*(6), 989–999. https://doi.org/10.1016/S0896-6273(03)00149-1

Polsky, A., Mel, B. W., & Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience, 7*(6), 621–627. https://doi.org/10.1038/nn1253

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Qi, W., Han, J., Zhang, Y., & Bai, L. (2017). Saliency detection via Boolean and foreground in a dynamic Bayesian framework. *The Visual Computer, 33*(2), 209–220. https://doi.org/10.1007/s00371-015-1176-x

Ray, S., & Maunsell, J. H. R. (2015). Do gamma oscillations play a role in cerebral cortex? *Trends in Cognitive Science*, *19*(2), 78–85. https://doi.org/10.1016/j.tics.2014.12.002

Regehr, W. G., Carey, M. R., & Best, A. R. (2009). Activity-dependent regulation of synapses by retrograde messengers. *Neuron*, *63*(2), 154–170. https://doi.org/10.1016/j.neuron.2009.06.021

Richard, A. M., Lee, H., & Vecera, S. P. (2008). Attentional spreading in object-based attention. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 842–853. https://doi.org/10.1037/0096-1523.34.4.842

Rideaux, R., Badcock, D. R., Johnston, A., & Edwards, M. (2016). Temporal synchrony is an effective cue for grouping and segmentation in the absence of form cues. *Journal of Vision, 16(11),* 23. https://doi.org/10.1167/16.11.23

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025. https://doi.org/10.1038/14819

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*, 203–227. https://doi.org/10.1146/annurev.neuro.29.051605.112939

Roelfsema, P. R., & de Lange, F. P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Annual Review of Vision Science*, *2*(1), 131–151. https://doi.org/10.1146/annurev-vision-111815-114443

Rutishauser, U., & Douglas, R. J. (2009). State-dependent computation using coupled recurrent networks. *Neural Computation*, *21*(2), 478–509. https://doi.org/10.1162/neco.2008.03-08-734

Rutishauser, U., Douglas, R. J., & Slotine, J.-J. (2011). Collective stability of networks of winner-take-all circuits. *Neural computation, 23,* 735–773. https://doi.org/10.1162/NECO-a-00091

Rutishauser, U., Slotine, J.-J., & Douglas, R. J. (2015). Computation in dynamically bounded asymmetric systems. *PLoS Computational Biology, 11*(1), e1004039. https://doi.org/10.1371/journal.pcbi.1004039

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5*(7), 631–632. https://doi.org/10.1038/nn876

Saenz, M., Buracas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research, 43*(6), 629–637. https://doi.org/ff246v

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Experimental Brain Research*, *77*(1), 23–30. https://doi.org/10.1007/BF00250563

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1–2), 1–46. https://doi.org/10.1016/S0010-0277(00)00152-9

Scimeca, J. M., & Franconeri, S. L. (2015). Selecting and tracking multiple objects. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(2), 109–118. https://doi.org/10.1002/wcs.1328

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, *55*(2), 301–312. https://doi.org/fs78wc

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *The Journal of Neuroscience*, *30*(9), 3531–3543. https://doi.org/bsk486

Spratling, M. W. (2011). A single functional model accounts for the distinct properties of suppression in cortical area V1. *Vision Research*, *51(6)*, 563–576. https://doi.org/10.1016/j.visres.2011.01.017

Spruston, N. (2008). Pyramidal neurons: Dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, *9*(3), 206–221. https://doi.org/10.1038/nrn2286

Tao, H. W., & Poo, M. (2001). Retrograde signaling at central synapses. *Proceedings of the National Academy of Sciences, 98*(20), 11009–11015. https://doi.org/bcj9v4

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135*(2), 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1628), 20130055. https://doi.org/10.1098/rstb.2013.0055

Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Tsui, J. M. G., Hunter, J. N., Born, R. T., & Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. *Journal of Neurophysiology*, *103*(6), 3123–3138. https://doi.org/10.1152/jn.00654.2009

Usher, M., & Cohen, J. D. (1999). Short term memory and selection processes in a frontal-lobe model. In D. Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience* (pp. 78–91). Springer-Verlag.

Vatterott, D. B., & Vecera, S. P. (2015). The attentional window configures to object and surface boundaries. *Visual Cognition, 23*(5), 561–576. https://doi.org/gh3q

Wang, D. L. (1999). Object selection based on oscillatory correlation. *Neural Networks, 12*(4), 579–592. doi: 10.1016/S0893-6080(99)00028-3

Wannig, A., Stanisor, L., & Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience, 18*(14), 1243–1244. https://doi.org/10.1038/nn.2910

Wei, D. S., Mei, Y. A., Bagal, A., Kao, J. P., Thompson, S. M., & Tang, C. M. (2001). Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science*, *293*(5538), 2272–2275. https://doi.org/fmn6z8

Yu, A. J., Giese, M. A., & Poggio, T. A. (2002). Biophysically plausible implementations of the maximum operation. *Neural Computation*, *14*(12), 2857–2881.

Yu, D., & Franconeri, S. L. (2015). Similarity grouping as feature-based selection. *Visual Cognition, 23*(7), 843–847. https://doi.org/10.1080/13506285.2015.1093234

Zhang, J., & Sclaroff, S. (2016). Exploiting surroundedness for saliency detection: A Boolean map approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(5), 889–902. https://doi.org/10.1109/tpami.2015.2473844

Zilberter, Y. (2000). Dendritic release of glutamate suppresses synaptic inhibition of pyramidal neurons in rat neocortex. *The Journal of Physiology, 528*(Pt 3), 489–496. https://doi.org/10.1111/j.1469-7793.2000.00489.x

Zilberter Y., Harkany, T., & Holmgren, C. D. (2005). Dendritic release of retrograde messengers controls synaptic transmission in local neocortical networks. *The Neuroscientist, 11*(4), 334–344. https://doi.org/10.1177/1073858405275827

Zilberter Y., Kaiser, K. M., & Sakmann, B. (1999). Dendritic GABA release depresses excitatory transmission between layer 2/3 pyramidal and bitufted neurons in rat neocortex. *Neuron, 24*(4), 979–988. https://doi.org/10.1016/S0896-6273(00)81044-2

APPENDIX B

Neural Dynamics of Spreading Attentional Labels in Mental Contour Tracing

Marić, M., & Domijan, D. (2019). Neural dynamics of spreading attentional labels in mental contour tracing. *Neural Networks, 119*, 113–138. https://doi.org/10.1016/j.neunet.2019.07.016

ABSTRACT

Behavioral and neural data suggest that visual attention spreads along contour segments to bind them into a unified object representation. Such attentional labeling segregates the target contour from distractors in a process known as mental contour tracing. A recurrent competitive map is developed to simulate the dynamics of mental contour tracing. In the model, local excitation opposes global inhibition and enables enhanced activity to propagate on the path offered by the contour. The extent of local excitatory interactions is modulated by the output of the multi-scale contour detection network, which constrains the speed of activity spreading in a scale-dependent manner. Furthermore, an L-junction detection network enables tracing to switch direction at the L-junctions, but not at the X- or T-junctions, thereby preventing spillover to a distractor contour. Computer simulations reveal that the model exhibits a monotonic increase in tracing time as a function of the distance to be traced. Also, the speed of tracing increases with decreasing proximity to the distractor contour and with the reduced curvature of the contours. The proposed model demonstrated how an elaborated version of the winner-takes-all network can implement a complex cognitive operation such as contour tracing.

Keywords: contour tracing; object-based attention; perceptual grouping; recurrent competitive map; winner-takes-all

1. INTRODUCTION

Vision starts with the parallel registration of features such as color, shape, or motion in dedicated processing streams (Lennie, 1998). The output of feature detectors is further elaborated by a set of Gestalt grouping rules to form a spatial representation of perceptual belongingness among objects and of figure-ground relationships (Wagemanas, 2014). Although the perceptual organization of a scene according to Gestalt rules involves sophisticated computations, it is not sufficient to extract all information that is of interest to the observer (Tsotsos et al., 2018). For instance, the perception of spatial relations between distal image parts, such as whether they are connected or whether they lie inside or outside of the same bounding surface, cannot be computed by any type of spatially limited feature detectors (Minsky & Papert, 1988). Figure 1 illustrates this fact. Whether the patterns presented in Figure 1A and 1B contain one or two black spirals is not immediately apparent.

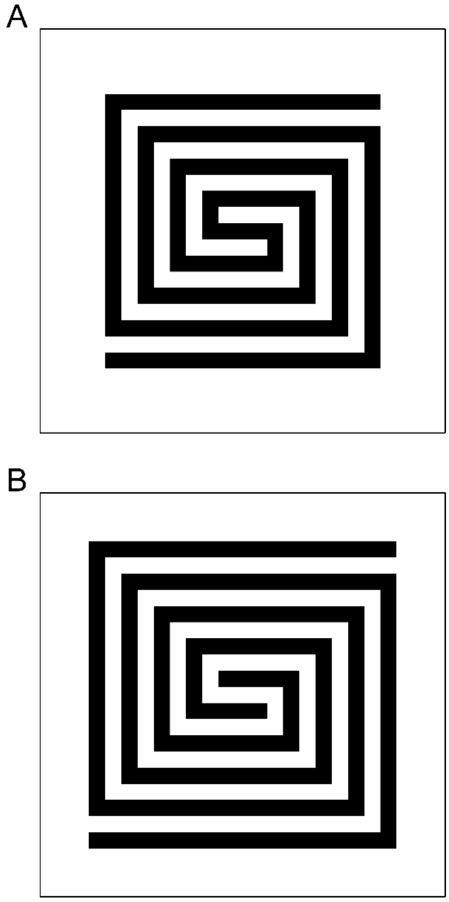


Figure 1. *The spiral problem. The task is to determine whether one (A) or two black spirals (B) exist in the image. The problem can be solved by mentally tracing the contour. This is an example of a visual routine.*

Ullman (1984, 1996) suggested that human observers comprehend spatial relations by applying visual routines on the representation offered by early vision. Visual routines refer to a set of cognitive operations that engage attention to bind together parts of the scene that remained ungrouped by the early visual representation. Attention labels distant features to render their spatial relations explicit. In a similar vein, Roelfsema and Houtkamp (2011) distinguished between *base grouping*, which depends on a fast extraction of image features and their conjunctions, and *incremental grouping*, which involves the engagement of slow, serial labeling of image elements that belong to the same perceptual group. Incremental grouping relies on object-based attention to highlight the representation of one perceptual group in an input image composed of many competing groups. Incremental grouping requires the establishment of dynamic links between neurons encoding features of the attended object and, at the same time, disabling connections to neurons encoding an unattended object.

Mental contour tracing is an example of a visual routine or incremental grouping process that has been extensively studied at both psychological and neural levels. It is engaged when we attempt to determine whether two image regions are connected, as illustrated in Figure 1. The detection of connectedness is important because connected image parts are likely to belong to the same objects, whereas disconnected parts usually belong to different objects (Roelfsema & Singer, 1998). In the laboratory, contour tracing is studied by a task where observers are required to determine whether two dots lie on the same contour in a pattern consisting of two (or more) intermingled contours. A typical finding is that the time it takes to provide an answer increases monotonically, but not linearly, with the distance between dots on the contour. The key factor determining the speed of tracing is the distance on the contour, and not the Euclidean distance between the dots, which is kept constant (Jolicoeur et al., 1986; Pringle & Egeth, 1988). Furthermore, tracing exhibits scale invariance since the absolute size of the contour does not influence the speed of tracing. This suggests the involvement of multiple spatial scales (Jolicoeur & Ingleton, 1991).

To isolate relevant factors contributing to the dynamics of tracing, Jolicoeur et al. (1991) devised simple stimuli consisting of a set of parallel straight lines or parallel curved lines. An example of the type of stimuli they used is depicted in Figure 2. Jolicoeur et al. (1991) systematically varied the distance between the target and distractor contours (Figure 2A–D) and the amount of curvature in the curved contours (Figure 2E–H) to measure their impact on the dynamics of tracing. Their results revealed that (a) the tracing time increases monotonically and roughly linearly with the length of the contour, (b) the tracing speed decreases with decreased spacing between the target and distractor contours, and (c) the tracing speed decreases with increased contour curvature. These findings help to explain why tracing times were not a linear function of the distance in studies employing two contours that wiggle around each other. In such stimuli, the distance between the target and distractor contours, as well as their curvature, vary considerably along the path that needs to be traced. Therefore, we will focus on the results of Jolicoeur et al.'s (1991) study in our modeling efforts.

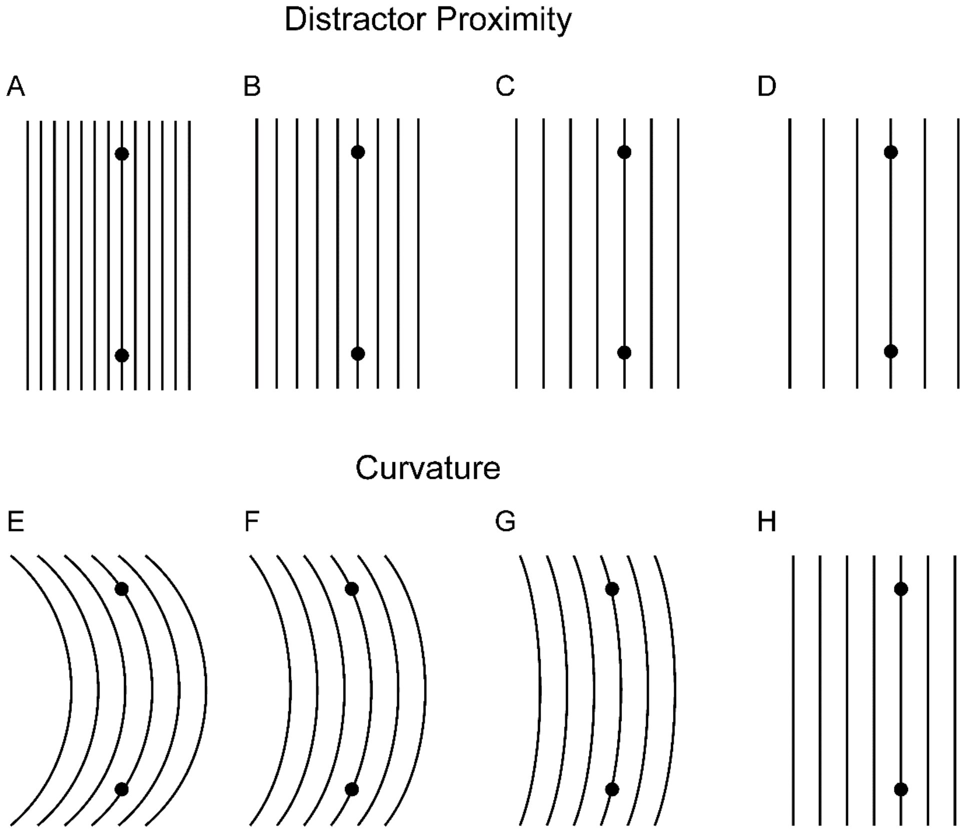


Figure 2. *The stimuli similar to those used by Jolicoeur et al. (1991) to examine the effect of distractor proximity (top row) and contour curvature (bottom row) on the dynamics of contour tracing. The task is to determine whether two dots lie on the same contour. The tracing time reduces as the distance between the target and distractor contours increases from very small (A), across small (B) and medium (C), to large (D). Also, when the proximity between contours is kept constant, the tracing time reduces as a function of contour curvature from large (E), across medium (F) and small (G) curvatures, to a straight contour (H).*

Several studies examined the question of whether attention moves or spreads along the contour. According to the zoom lens model, attention makes discrete jumps from one part of the contour to the next. The size of jumps is flexibly adjusted to avoid making mistakes

(McCormick & Jolicoeur, 1991, 1994). Therefore, the size of the jump is smaller when the target contour is near the distractor, and it increases when the distractor contour is further away. Crundall et al. (2008) provided further support for this model by demonstrating that participants do not notice changes that occur near the beginning of the contour, when the tracing operator has sufficient time to move away from the starting point.

On the other hand, three studies (Houtkamp et al., 2003; Roelfsema et al., 2010; Scholte et al., 2001) found that tracing involves highlighting all elements of the same contour. These results imply that tracing operates similarly to object-based attention: it selects all spatial locations occupied by the same object. In other words, tracing creates a visual representation where a grouped array of locations is selected (Cosman & Vecera, 2012; Hollingworth et al., 2012; Vatterott & Vecera, 2015).

Neural recordings in the monkey primary visual cortex (V1) suggest that contour tracing is associated with elevated firing rates in neurons whose receptive fields fall on the target contour, relative to neurons whose receptive fields fall on the distractor contour (Roelfsema et al., 1998). Next, it was found that firing rate modulation occurs earlier for neurons located near the start of the tracing process (fixation point) and later for neurons located further away along the contour. Importantly, the response enhancement for neurons encoding early segments of the contour remained approximately constant during the whole trial, thus providing direct support for the idea that attention spreads rather than moves along the contour (Roelfsema et al., 2003). Moreover, the timing of the response enhancement on neurons encoding a distal contour segment depends on how close the target and distractor contour are placed (Pooresmaeili & Roelfsema, 2014). If there is a small gap between proximal segments of the target and distractor contours, then the response enhancement on the distal segment of the target contour is delayed relative to the stimulus with a wider gap. This is consistent with the behavioral findings on the effect of contour spacing on the speed of tracing (Jolicoeur et al., 1991).

Wannig et al. (2011) found that enhanced activity is initiated by the external cue and automatically spreads from the cued to the neighboring neurons if they share the same feature selectivity, namely color or orientation. Interestingly, attentional modulation in the contour tracing task was not observed in all tested neurons. About half of the neurons were not affected by the attention at all, and they were labeled as N-neurons – as opposed to A-neurons, which exhibited response enhancement (Pooresmaeili et al., 2010). Finally, it should be noted that the reviewed studies were not able to discern whether the source of the tracing signal arises from the horizontal connections within the V1 or via feedback connections from the extrastriate cortex. Roelfsema and Houtkamp (2011) proposed that contour tracing is a consequence of the interactions within and between cortical areas V1, V2, and V4 organized in a dynamic processing hierarchy termed a *growth cone*. The size of the cone is dynamically adjusted depending on the stimulus conditions. When the target and distractor contours are far apart, a larger cone is activated that encompasses higher hierarchical levels containing neurons with larger receptive sizes. Contour tracing consequently advances faster along the contour (see also Jeurissen et al., 2016; Pooresmaeili & Roelfsema, 2014).

The aim of the present study is to develop a neurocomputational account of the dynamics of mental contour tracing consistent with the reviewed behavioral and neural data. The model rests upon a feature-based winner-takes-all (F-WTA) network, recently proposed by Marić and Domijan (2018). The F-WTA network is a recurrent competitive map with local interactions between excitatory units and global inhibition mediated by a single inhibitory unit. It is capable of simultaneous selection of many winners based on top-down guidance. We have demonstrated how to embed the F-WTA network into a larger neural architecture incorporating multi-scale contour and L-junction detection networks. The output of the contour and L-junction detection is used to guide the lateral excitatory interactions within the F-WTA network. In this way, enhanced activity in the F-WTA network spreads along the target contour as fast as possible without making mistakes – that is, without activity spillover to the distractor contour.

2. MODEL DESCRIPTION

2.1. Model Overview

The neural model of contour tracing consists of three components: The F-WTA network, the contour detection network (CDN), and the L-junction detection network (LDN). In this chapter, the components are informally described first to provide an understanding of how they contribute to the tracing. In the second part, the formal specification of each model component is provided.

2.1.1. Feature-based Winner-Takes-All Network

Computational models of visual attention rely on competitive neural networks to find the locations of greatest importance or saliency (Itti & Koch, 2001; Tsotsos, 2011). An extreme form of competition is implemented in the winner-takes-all (WTA) network, which selects only the most active location in the input image (Koch & Ullman, 1985; Yuille & Geiger, 2003). A physiologically realistic model of cortical competition is a WTA network with asymmetric connectivity. It consists of an array of excitatory nodes that are reciprocally connected to a single inhibitory interneuron (Rutishauser & Douglas, 2009; Rutishauser et al., 2011). Such an anatomical arrangement creates a global *blanket of inhibition* to excitatory nodes (Fino et al., 2013). However, the WTA network quickly loses its capacity to represent winners when the number of nodes receiving maximal input increases (Haarmann & Usher, 2001; Usher & Cohen, 1999). To prevent this decline, it is important to control excessive inhibition in a situation where many winners need to be selected simultaneously, as required by feature- and object-based attention.

We recently developed a model of the asymmetric cooperative-competitive network depicted in Figure 3. The F-WTA network is capable of simultaneous selection of an arbitrary number of locations based on their shared feature value (Marić & Domijan, 2018). It incorporates dendrites (Häusser & Mel, 2003; London & Häusser, 2005; Mel, 2016) and synapses (Abbott & Regehr, 2004; Regehr et al., 2009) as independent computational units. Model dendrites mediate recurrent excitation between adjacent neurons, and their nonlinear output function prevents unbounded growth of neural activity. Furthermore, the retrograde inhibition of synaptic transmission enables the inhibitory interneuron to compute the maximum function instead of the sum over its input signals (Kaski & Kohonen, 1994). The inhibitory interneuron sets a dynamic quenching threshold (QT) for the excitatory nodes (Grossberg, 1973). It inhibits all nodes that are below the QT and spares all nodes that are above the QT. Importantly, the QT tracks the activity of the winning nodes when it changes because of the change in the input amplitude. In this way, the network performs a state-dependent computation driven by external inputs (Rutishauser & Douglas, 2009). A detailed description of the way in which the QT relates to the network activity is provided in the Appendix.

Retrograde inhibition on the excitatory nodes also ensures that winners do not inhibit one another, thereby enabling simultaneous selection of the arbitrary number of winners if they share the same input amplitude. This leads to a form of winners-share-all solution rather than to a more restrictive selection of a single node (Fukai & Tanaka, 1997). With these properties of the F-WTA network, we were able to simulate many important findings regarding the formation of feature-based spatial maps and their elaboration by the set operations of intersection and union (Huang & Pashler, 2007).

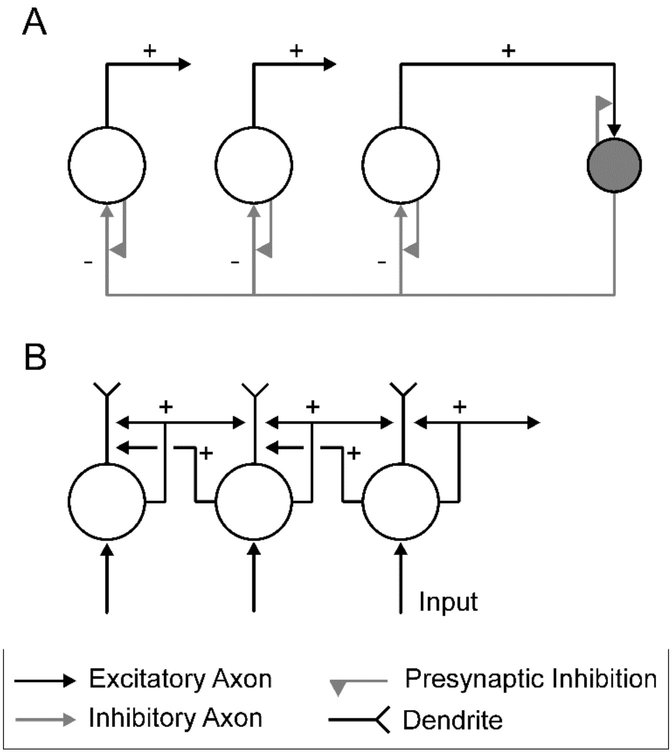


Figure 3. *A feature-based winner-takes-all (F-WTA) network. A set of excitatory nodes (white disks) are reciprocally connected to a single inhibitory node, depicted as a gray disk (A). Both types of nodes release retrograde messengers that inhibit the transmitter release from the presynaptic terminal in proportion to the node's activity. This retrograde signaling creates a negative feedback loop that dynamically regulates the synaptic transmission between excitatory and inhibitory nodes. In addition, each excitatory node receives direct feedforward input and recurrent input from its dendrite, which is modeled as a separate compartment (B). The dendrite receives the node's self-excitation and recurrent collaterals from the neighboring nodes.*

2.1.2. Contour Detection Network

In the original formulation of the F-WTA network, lateral excitatory connections among nodes are isotropic and restricted to nearest neighbors. Such a connectivity scheme results in slow pixel-by-pixel activity spreading in all directions. Here, we demonstrate how to enrich the lateral connectivity of the F-WTA network in order to simulate properties of mental contour tracing. First, we assume that separate F-WTA networks exist for each orientation. Excitation within each network is thus restricted to the direction of preferred orientation only. In this way, activity spreading in one orientation will not jump to other orientations at the contour junction. For computational simplicity, we employed only two orientations: horizontal and vertical. Next, to account for a variable speed of contour tracing, we propose that the F-WTA nodes do not interact directly with one another. Rather, they are engaged in a feedback loop with the same-oriented multi-scale CDN. In particular, lateral excitatory interactions within the F-WTA network are multiplicatively gated by the output of the CDN (Figure 4).

We adopted a simple model of contour detection with two orientations (horizontal and vertical) and four spatial scales (tiny, small, medium, and large, denoted as 1, 2, 3, and 4, respectively). The input image is convolved with a set of multi-scale, even-symmetric Gabor filters to simulate oriented receptive fields known to exist in the visual cortex (Daugman, 1980; Marčelja, 1980). The output of the convolution is provided by *c*-units. In addition, *a*-units receive scale-dependent feedback projections from the F-WTA network with the same orientation preference. Importantly, the lateral extent of the feedback projections to the *a*-units scales with the size of the receptive field of the corresponding *c*-units, as illustrated in Figure 4. The output of the *a*-units is multiplicatively gated by the *c*-units before it is fed back to the F-WTA network. In this way, the *c*-units control the size of the integration zone over which the F-WTA network samples the activity of its neighbors via *a*-units. The same connectivity pattern is replicated in the vertical F-WTA network, except that the lateral interactions extend in the vertical instead of horizontal direction. The distinction between the *a*-units, which receive attentional signals from the F-WTA network, and the *c*-units, which are unaffected by attention, is reminiscent of the distinction between the A- and N-units found in the visual cortex (Pooresmaeili et al., 2010). Finally, it should be noted that we designed the model of *c*-units to be as simple as possible. However, it is not too difficult to replace this model with more sophisticated models of contextual interactions in the visual cortex that are capable of robust contour detection (Grossberg et al., 1997; Hansen & Neumann, 2004; Ursino & La Cara, 2004).

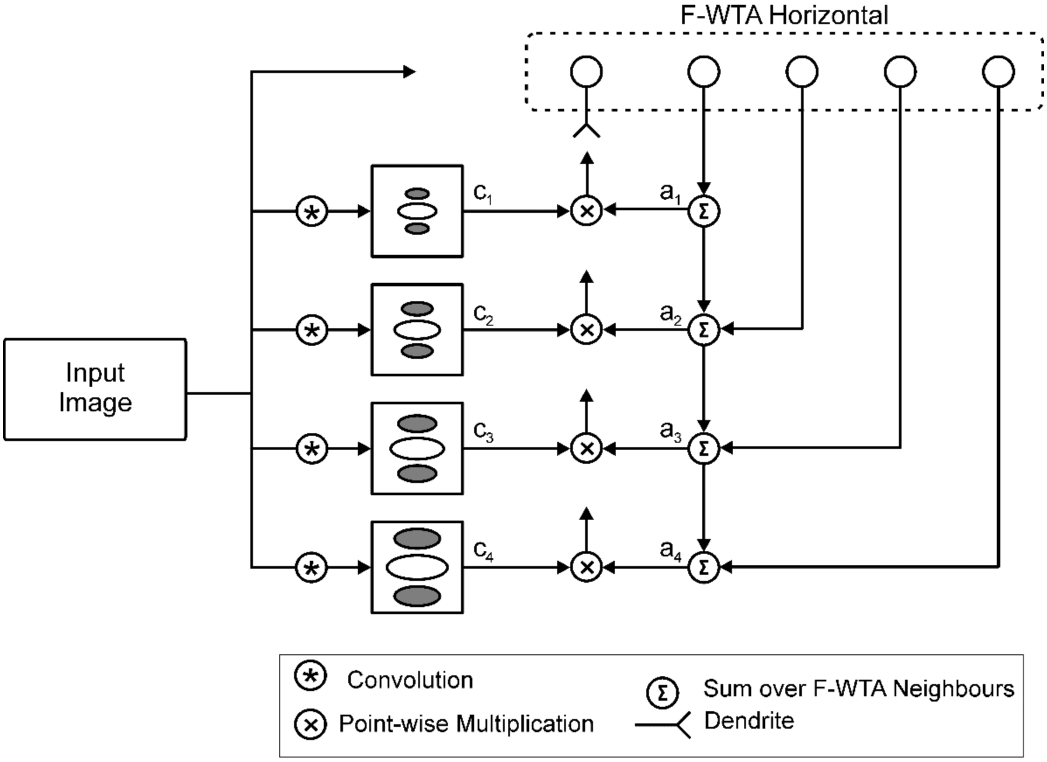


Figure 4. *A multi-scale neural model of contour tracing. In the model, the contour detection network (CDN) controls the scale-dependent excitatory feedback loop of the F-WTA network. For simplicity, we depict the connections within the horizontal F-WTA network only. In the CDN, the input image is convolved with a set of Gabor filters of four different sizes to produce feedforward output (c_1, c_2, c_3, c_4). The filter's output multiplicatively gates the feedback signals arising from the F-WTA network (a_1, a_2, a_3, a_4). The size of the window over which the F-WTA activity is sampled in the a-units correlates with the size of the Gabor filter in the c-units. An externally supplied spatial cue drives the activity spreading by temporarily increasing the activity of one of the nodes in the network.*

It is important to emphasize that the F-WTA network, not the CDN, is a carrier of the activity spreading related to attention. The role of the CDN is to open or close the feedback loops between neighboring nodes in the F-WTA in a scale-dependent manner. The CDN modulates the size of lateral interactions between same-oriented nodes in the F-WTA network. Figure 5 illustrates how the presence or absence of CDN activity at a certain scale widens or narrows the size of the integration window within the F-WTA network. The CDN network serves as a cognitive blackboard on which the F-WTA network writes and reads its activity in order to gauge the size of the next tracing step (Roelfsema & de Lange, 2016). At each location in the map, the F-WTA network broadcasts its message to all scales of the CDN and waits to observe which scale will respond back. As more scales are activated to a suprathreshold level in the CDN, more distant F-WTA nodes may exchange excitation and thus speeds up the progress of activity spreading. Further details about this process are provided in Sections 2.2.1. and 2.2.2.

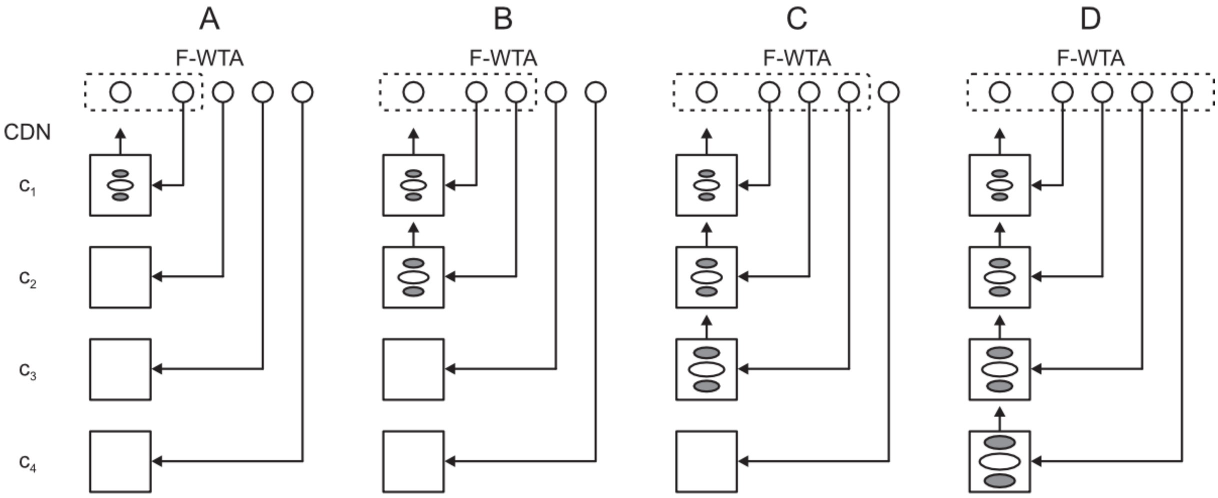


Figure 5. *Scale-dependent interactions between the CDN and the F-WTA network. An empty box indicates subthreshold activity at the given scale. As more scales are activated in parallel (from one to four scales as we move from A to D), the node in the F-WTA network receives excitation from a wider neighborhood (denoted by a dashed rectangle), thereby leading to a faster contour tracing.*

2.1.3. L-junction Detection Network

Interaction between the F-WTA and the CDN helps to explain how contour tracing evolves within the single orientation. However, there is also a need for cross-orientation interactions. Tracing can switch between orientations at corners or L-junctions where the same contour is bent. On the other hand, tracing should not switch orientation at T- or X-junctions, because they occur at the intersection of two distinct contours. Furthermore, X- and T-junctions are particularly problematic for contour tracing because there is a danger of confusing the target with the distractor contour (Ullman, 1984). Therefore, the model should allow activity spreading across orientations at the L-junctions but not at other types of junctions.

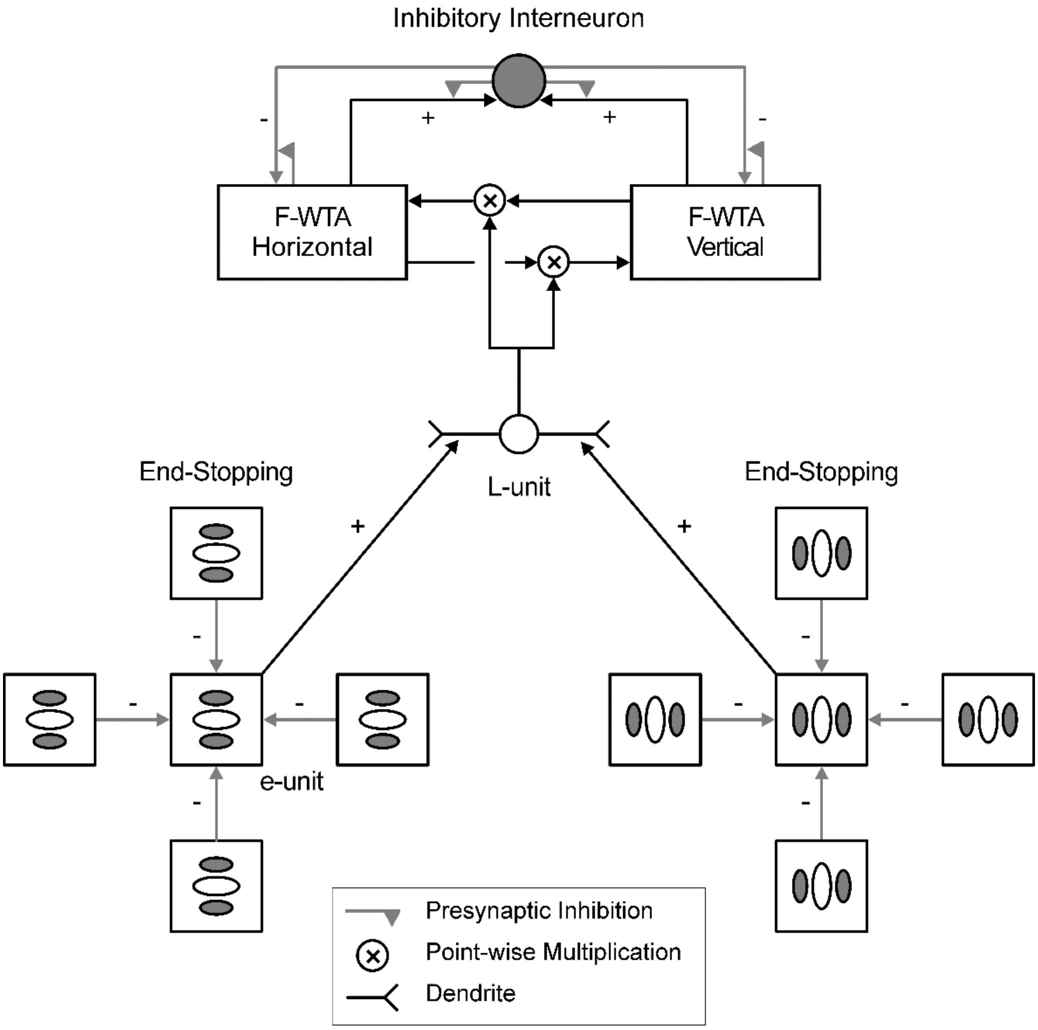


Figure 6. *The L-junction detection network (LDN) controls excitatory interactions between horizontal and vertical F-WTA networks. Lateral inhibition among the output of the same-oriented c-units creates an end-stopping response in the e-unit. The l-unit detects the superposition of perpendicular e-units. Moreover, the l-unit is activated only when both of its two dendrites are active. Next, the l-unit multiplicatively gates recurrent collaterals connecting two F-WTA networks. In addition, the F-WTA networks interact with the joint inhibitory interneuron that mediates global lateral inhibition.*

To this end, we introduce the third component of the model, that is, the LDN (Figure 6). It is composed of *e*-units and *l*-units. The *e*-units exhibit end-stopping, that is, enhanced response at contour ends and corners. End-stopping is a biophysically plausible mechanism to create various types of junction detectors (Thielscher & Neumann, 2008). The *e*-units receive input from the same-oriented *c*-units. Lateral inhibition among the same-oriented receptive fields across neighboring positions leads to a stronger response at the line end and a weaker response elsewhere. Next, the *l*-unit detects superposition of the end-stopping activity. The *l*-unit has two dendrites, each receiving input from a distinct *e*-unit at the same location. The threshold of the *l*-unit is set in a way to ensure that it operates as an AND gate. It is activated only when it receives suprathreshold signals from both *e*-units. In this way, the *l*-unit selectively responds to the corner but not to a single line end. In other words, the *l*-unit will selectively respond to an L-junction but not to a T- or an X-junction. The T-junction will activate only a single *e*-unit, and the X-junction will not activate *e*-units at all. Finally, the output of the *l*-unit establishes the link between the horizontal and vertical F-WTA networks via the same multiplicative gating as that used in the CDN. The link is active only at the L-junctions, thereby enabling activity spreading to change direction when those junctions are encountered (Figure 7).

Here, we assume that the *e*- and *l*-units exist at the smallest spatial scale only, although this is not critical for the model's operation. Empirical support for the corner detectors was provided by Anzai et al. (2007), who found evidence that some of the neurons in the V2 area encode a combination of orientations rather than a single orientation. Those neurons primarily respond to a combination of orientations that are 90º apart, consistent with their role as L-junction or corner detectors. Finally, it should be noted that the LDN could be generalized to a model with many orientations. The only requirement is that the *l*-unit receives input from each *e*-unit on a separate dendrite. In that case, the *l*-unit will be sensitive to any acute or obtuse angle formed by two non-colinear contour segments. In such a model, there should be as many F-WTA networks as there are orientations. In addition, they should all exchange gated excitation among one another at the same location. Activity spreading may then proceed in directions that are consistent with the orientation of contour segments joining at the L-junction.

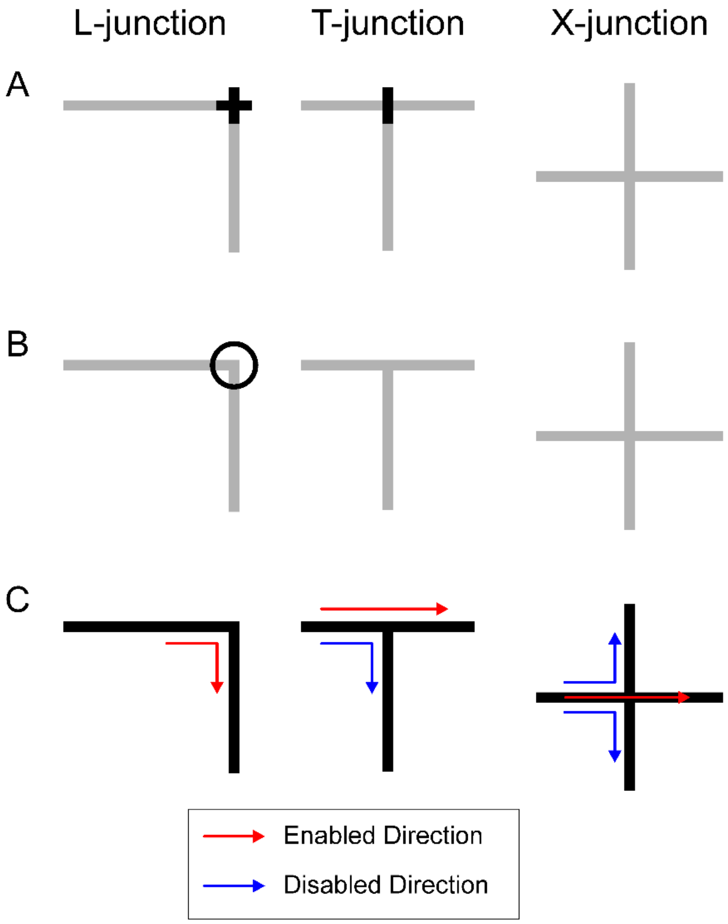


Figure 7. *The differential responses of the LDN to the L-, T-, and X-junctions. End-stopping or e-units display an enhanced response at line ends (A). On the one hand, at the L- junction, two line ends exist, so both horizontal and vertical e-units are strongly activated (short black lines). On the other hand, at the T-junction, there is only one active e-unit. At the X-junction, there are no line ends and consequently no suprathreshold activity among e-units. As a consequence, the l-unit (black circle) shows suprathreshold activity only at the L-junction (B). Finally, the suprathreshold activation of the l-unit enables activity spreading in the F-WTA network to switch direction at the L-junction. This is not possible at the T- or X-junctions because there is no active l-unit (C).*

2.2. Formal Description

The mathematical description of the model is provided in the next four sections.

2.2.1. Feature-based Winner-Takes-All Network

The dynamics of the F-WTA network is defined using differential equations. They specify the time-varying activity (or firing rate) of a population of excitatory nodes x_{ijk} and a single inhibitory node *y*. The activity of the excitatory node x_{ijk} at position (*i*, *j*) and orientation index *k* obeys the following equation

$$\tau_x \frac{dx_{ijk}}{dt} + x_{ijk} = f\left[I_{ij} + J_{ij} + \alpha g\left(x_{ijk} + C_{ijk} + L_{ijk} - T_d\right) - \beta_1 h\left[y - x_{ijk} - T_y\right]\right]. \tag{1}$$

The activity of the inhibitory interneuron *y* obeys the following equation

$$\tau_y \frac{dy}{dt} + y = f\left[\beta_2 \sum_{i,j,k} h\left[x_{ijk} - y - T_x\right]\right]. \tag{2}$$

Parameters τ_x and τ_y control how quickly excitatory and inhibitory nodes respond to their inputs, respectively. The membrane potential of each node is transformed into an instantaneous firing rate by the somatic activation function *f*[*u*] defined by

$$f\left[u\right] = h\left[u\right] = \max\left(u, 0\right). \tag{3}$$

The rectification is also used to describe the output of synaptic interactions *h*[*u*] as explained below. The excitatory node receives constant bottom-up input I_{ij} and time-varying top-down cue J_{ij}, which is switched off by default. This cue is temporarily switched on at the location where the tracing should begin. When and where the top-down cue is switched on is decided in a task-dependent manner by some unspecified process outside of the F-WTA network. Tsotsos and Kruijne (2014) described network components that can implement complex cognitive programs and that can serve as a source of the top-down signals. In addition to the bottom-up and top-down input, the soma of the excitatory node also receives input from its dendrite, which operates as an independent computational unit with its own activation function. Dendritic output *g(u)* is given by the piecewise-linear function of the form

$$g\left(u\right) = \begin{cases} 0, & if \quad u \le 0, \\ u, & if \quad 0 < u < 1, \\ 1, & if \quad u \ge 1, \end{cases} \tag{4}$$

and T_d is a threshold for dendritic activation. In Equation (1), parameter α denotes the impact of the dendritic compartment on the soma. A dendrite receives self-recurrent collateral x_i, gated input from the CDN denoted by C_{ijk}, and input from the LDN denoted as L_{ijk}. They will be described in Sections 2.2.2. and 2.2.3., respectively. This means that all recurrent excitatory connections arrive on a single dendrite. Sigmoid nonlinearity, given by Equation (4), bounds from above the total amount of excitation a node can receive. In this way, the dendrite prevents runaway excitation and allows the node to remain sensitive to input fluctuations and to contextual interactions arising from the local neighborhood. This is a key model property necessary to achieve activity spreading.

The excitatory node also receives lateral inhibition via the inhibitory interneuron y. However, lateral inhibition is counteracted by retrograde signaling from the excitatory node to the presynaptic terminal. In Equation (1), the term $-h[y - x_{ijk} - T_y]$ describes the interaction that takes place at the presynaptic terminal that delivers inhibition from interneuron y to excitatory node x_{ijk}. Rectification $h[u]$ ensures that the presynaptic terminal will release its inhibitory transmitter only when the activity of node y exceeds the inhibitory retrograde signal from the postsynaptic node $-x_i$ and the threshold for presynaptic activation denoted as T_y. The strength of the inhibition is determined by parameter β_1. A detailed biophysical justification for such synaptic interactions was provided in Marić and Domijan (2018).

In a similar vein, in Equation (2), the term $-h[x_{ijk} - y - T_x]$ describes an action of the retrograde signal that is released from inhibitory interneuron y on the presynaptic terminal that delivers excitation from node x_{ijk}. Here, parameter T_x describes the threshold for the activation of the presynaptic terminal of the excitatory node, and β_2 determines the strength of the excitation. Importantly, the same inhibitory interneuron receives excitatory input from both F-WTA networks in order to coordinate tracing across orientations. However, it is possible to rearrange the model so that each F-WTA network has its own inhibitory interneuron. In this case, it is necessary to include a separate excitatory node, which computes a maximum within its own network and projects its output to the inhibitory interneurons in other F-WTA networks (Rutishauser et al., 2012). Also, we note that inhibitory interneuron y receives only excitatory input from the x nodes. Still, we applied the same activation function $f[u]$ on its total membrane potential in order to provide consistent descriptions of all model components.

2.2.2. Contour Detection Network

The total output of the CDN C_{ijk} to the F-WTA network at location (i, j) and orientation k is given by the sum over all four spatial scales $s = \{1, 2, 3, 4\}$ corresponding to the tiny, small, medium, and large scale, respectively

$$C_{ijk} = \sum_{s} \omega_s a_{ijks} c_{ijks} \ . \tag{5}$$

Parameters ω_s denote scale-dependent synaptic weights controlling the impact of the CDN on the dendrites of the F-WTA network. The feedback projection from this network to the a-unit at scale s is given by

$$a_{ijks} = f\left[\sum_{p,q \in N_{ijks}} x_{i+p,\, j+q,k} \right] . \tag{6}$$

Although the a-unit receives only excitatory input from the F-WTA network, its membrane potential passes through activation function $f[u]$ for completeness. The set of indices N_{ijks} describe the window over which the a-unit samples back-projected activity from the x nodes (Figure 4). It is defined separately for a small scale (s = 1) where

$$N_{ijk1} = \left\{ (i, j) : -1 \le i \le 1, -1 \le j \le 1 \right\} \tag{7}$$

across all orientations k. This means that the a_1-unit samples the activity of the F-WTA node at location (i, j) and its eight nearest neighbors. Next, a set of indices for larger scales (s > 1), is given by

$$N_{ij1s} = \left\{ (i, j) : 0, -s \le j \le s \right\} \tag{8}$$

for a horizontal orientation, and by

$$N_{ij2s} = \left\{ (i, j) : -s \le i \le s, 0 \right\} \tag{9}$$

for a vertical orientation.

The activation of the *c*-unit at location {*i, j*} is obtained by the convolution of the input image with a set of Gabor filters of different orientations *k* and scales *s*

$$c_{ijks} = f\left[\sum_{p,q} G(p,q,k,s)\, I_{i+p,j+q} - T_s\right] \tag{10}$$

where *f*[*u*] is the activation function defined by Equation (3); the set of indices -15 < *p, q* < 15 defines the spatial window for convolution; T_s is a scale-dependent threshold; and

$$G(p,q,k,s) = \exp\left(-\frac{p'^2 + \gamma^2 q'^2}{2\delta_s^2}\right)\cos\left(2\pi\frac{p'}{\lambda_s} + \psi\right) \tag{11}$$

is an even-symmetric Gabor filter with orientation index *k*, aspect ratio *γ*, standard deviation δ_s, wavelength λ_s, and phase *ψ*. The coordinate transformation that rotates the filter to an angle θ_k is given by

$$p' = p\cos(\theta_k) + q\sin(\theta_k) \tag{12}$$

and

$$q' = -p\sin(\theta_k) + q\cos(\theta_k), \tag{13}$$

where $\theta_k = 2\pi k/K$, $k \in \{1,2\}$ and $K = 2$ is the total number of orientations. Each filter has three subregions that extend along the preferred axis: one excitatory subregion in the middle of the filter and two flanking inhibitory subregions.

The dynamics of the *a*- and *c*-units in the CDN were not explicitly modeled. Instead, we assume that they react to their respective input in an infinitely fast manner, and we inserted their instantaneous equilibrium activation into the Equation describing the dynamics of the F-WTA network. The same assumption is made for the LDN nodes to be described below.

2.2.3. L-junction Detection Network

The cross-orientation interaction between the F-WTA networks L_{ijk} at location (i, j) and orientation k is defined by

$$L_{ijk} = x_{ijk'1} \omega_L l_{ij} . \tag{14}$$

In Equation (14), *k'* denotes an orientation perpendicular to k, while ω_L denotes the synaptic weight controlling the strength of cross-orientation excitation, and the *l*-unit activity l_{ij} is given by

$$l_{ij} = f\left[\sum_k g\left(e_{ijk1}\right) - T_L \right] . \tag{15}$$

In Equation (15), *f[u]* is an activation function of the *l*-unit defined by Equation (3), *g(u)* is an activation function of its dendrites defined by Equation (4), and T_L is a threshold. Each dendrite receives input from the end-stopping unit of different orientation k. The activity of the end-stopping unit e_{ijk1} at the location (i, j), with orientation k, and at the smallest scale (s = 1) is given by

$$e_{ijk1} = f\left[\frac{\varepsilon c_{ijk1}}{1 + c_{ijk1} \sum\limits_{(p,q) \in E_{ij}} c_{i+p,j+q,k,1}} - T_E \right] . \tag{16}$$

Parameter ε controls the gain of the *e*-unit, and T_E is a threshold for its activation. The set E_{ij} of locations consists of four nearest neighbors

$$E_{ij} = \left\{ \left(i, j-1\right), \left(i-1, j\right), \left(i+1, j\right), \left(i, j+1\right) \right\} . \tag{17}$$

Equation (16) describes divisive inhibition by the same-oriented neighboring *c*-units. In the denominator, the summed input from neighboring *c*-units is multiplicatively gated by the activity of the target *c*-unit. In this way, the *e*-unit receiving weak excitation and strong lateral inhibition will exhibit a stronger response relative to the *e*-unit receiving strong excitation and strong lateral inhibition. The *e*-unit will consequently exhibit an enhanced response at line ends and a weaker response further away from the line ends.

2.2.4. Parameters and Simulation Method

For simplicity, the thresholds that control the activation of the excitatory and inhibitory nodes in the F-WTA network are set to zero, and they are omitted from the model description. The parameter values of the F-WTA networks were set as follows: τ_x = 40 ms, τ_y = 20 ms, α = 1, β_1 = 1, β_2 = 10, T_d = 0.2, T_x = 0.2, and T_y = 0.2. The parameter values of the Gabor filters in the CDN were set as follows: γ = 0.5, λ_1 = 3, λ_2 = 4, λ_3 = 5, λ_4 = 6, $\delta_1 = \lambda_1/3$, $\delta_2 = \lambda_2/3$, $\delta_3 = \lambda_3/3$, $\delta_4 = \lambda_4/3$, and ψ = 0. Furthermore, the scale-dependent thresholds for the activation of the *c*-units were set to: T_1 = 1.5, T_2 = 4, T_3 = 6, and T_4 = 8. The synaptic weights controlling the impact of the CDN on the dendrites of the F-WTA networks were set to ω_1 = 1 in the simulations restricted to the vertical orientation (3.1., 3.3., and 3.4.) and to ω_1 = 2 in all other simulations. All other weights were kept constant across all simulations, and they were set to ω_2 = 1, ω_3 = 1, and ω_4 = 1. Finally, the parameters of the LDN were set to ω_L = 2, T_L = 1.5, T_E = 1.2, and ε = 10.

The simulations were conducted by numerically integrating Equations (1) and (2) using Euler's method with a time step Δt = 0.05 ms. As a verification of the numerical stability of the integration, simulation 3.1. was rerun with Δt = 0.01 ms, and the results were identical to the original simulation. To illustrate the model's behavior, we run a set of computer simulations with stimuli such as those used in the empirical studies on mental contour tracing. The size of the input image was 30 × 30 in simulations 3.1. and 3.3.; 40 × 40 in simulations 3.2. and 3.4.; 19 × 20 in simulation 3.5.; and 25 × 25 in simulations 3.6. and 3.7. If the location was occupied by the contour, then we set I_{ij} = 1, and the background locations were set to I_{ij} = 0. When the cue was applied to a location (i, j), we set J_{ij} = 2, otherwise it was set to J_{ij} = 0. In all simulations, the cue was applied in the interval t = [200 ms, 400 ms]. Each image has been scaled so that the maximum activity and minimum activity are mapped to black and white, respectively. Intermediate values map linearly onto a gray scale.

3. RESULTS

3.1. The Effect of Distractor Proximity

The basic property of contour tracing is that the time to connect starting and ending points on the contour increases monotonically with the distance between these points. Moreover, when tracing occurs along straight lines, its speed is modulated by the proximity of the target and distractor contours (Jolicoeur et al., 1991). As the proximity is increased, the tracing becomes increasingly slower, even though the distance to be traced is kept constant. To demonstrate that the proposed model can account for these properties, we performed a simulation with the input image consisting of vertical lines separated horizontally by a variable number of pixels.

Figure 8 depicts four different spacings of vertical contours and the corresponding responses of all four scales of the *c*-units. Here, we illustrate the responses of the vertical *c*-units only because the responses of both the horizontal *c*-units and the *l*-units are subthreshold. As can be seen, the output of the vertical *c*-units correlates with the contour spacing. When the horizontal separation between contours was only one pixel wide (Figure 8A), the contours were encoded at the tiny scale or c_1-units only. The reason for this is scale-dependent flanking inhibition of the Gabor filters. At the small, medium, and large scales, neighboring contours fell within the inhibitory zone of the corresponding Gabor filters and reduced the activation of the c_2-, c_3-, and c_4-units. In addition, scale-dependent thresholds T_s make these weak activations subthreshold.

The exception to the inhibitory effect of neighboring contours is *c*-units encoding contours positioned at the edge of the grating. They receive weaker total inhibition and consequently survive thresholding even at larger scales. In other words, they receive inhibition only from one flank, while all other *c*-units receive inhibition from both flanks, thereby creating this edge effect. When the horizontal separation between contours was two pixels wide (Figure 8B), contours were encoded at tiny and small scales but not at the medium and large scales. In this case, the c_2-units are released from the flanking inhibition because neighboring contours are positioned farther apart relative to their inhibitory zones. Next, when the horizontal separation between contours was three pixels wide, the c_3-units were also released from inhibition in addition to the c_1- and c_2-units (Figure 8C). Finally, when the distance between contours was four pixels wide (Figure 8D), the *c*-units at all scales were released from flanking inhibition. All contours were consequently encoded at all four scales.

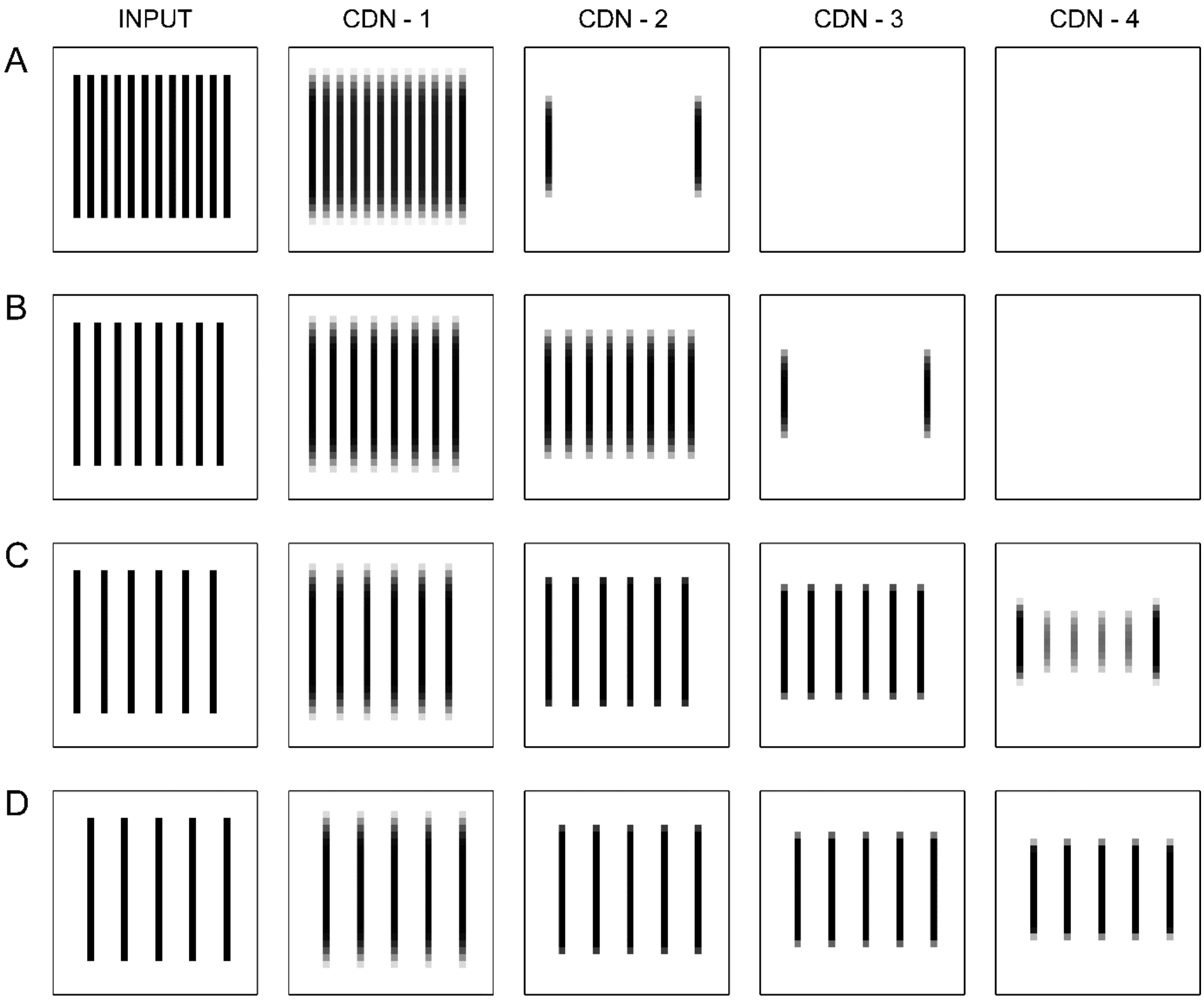


Figure 8. *Simulation of the effect of distractor proximity on the output of the vertical c-units at all four spatial scales s = {1, 2, 3, 4}. Input is a vertical grating with a one-pixel (A), two-pixel (B), three-pixel (C), or four-pixel (D) distance between contours.*

Figure 9 displays snapshots of the evolving F-WTA activity taken at five representative time points in response to four input configurations depicted in Figure 8. Each row depicts the F-WTA response to one stimulus condition: one-pixel (A), two-pixel (B), three-pixel (C) and four-pixel (D) wide separation between contours. We opted to display the network activity at five points in time that are of special interest in understanding network behavior. Those are the presentation of the spatial cue (*t* = 300 ms), removal of the cue and the start of activity spreading (*t* = 500 ms), spreading of activity enhancement along the target contour (*t* = 600 ms and t = 800 ms), and the end of the simulation (*t* = 1500 ms).

At the start of the simulation, all contours were selected together because they all share the same input amplitude (not shown). Next, the spatial cue was applied to the network at the starting location for tracing. Here, we arbitrarily chose a contour in the middle of the image as a target for tracing. Importantly, the cue is external to the network. It is not possible to generate tracing automatically within the network. This is consistent with the findings of Crundall, Cole, et al. (2008), who demonstrated that tracing is not an obligatory process that is automatically activated upon the stimulus presentation. Rather, it is a strategy that could be engaged by top-down signals depending on the current task demands. Interestingly, Crundall, Cole, et al. (2008) also found that when the tracing started, it could not be stopped, because they observed that tracing proceeds even to the contour segments that are irrelevant to the task. Moreover, Donovan et al. (2017) found that spatial cueing is a necessary condition to observe the effects of object-based attention.

After the cue was applied, the F-WTA network began to inhibit all non-cued locations (Figure 9, *t* = 300 ms). At this moment, the F-WTA network behaved as a standard WTA network that selects the node with the maximal input amplitude, that is, the cued location alone. The activity of all other nodes fell below the QT – they were inhibited and eventually completely removed from the representation. If the spatial cue is constantly supplied to the network, then it will remain in its steady state corresponding to the WTA solution. However, an interesting behavior in the F-WTA network occurred after the spatial cue was withdrawn at *t* = 400 ms (not shown). Instead of returning to the initial state where all contours were selected together, the network gradually enhanced and selected the nodes that were connected to the cued location (Figure 9, *t* = 500 – 1500 ms). This is a consequence of the lateral excitatory interactions mediated by the CDN, as schematically illustrated by Figure 5. Activity supplied by the CDN enabled the inactive neighboring nodes of the currently active node in the F-WTA network to cross the QT and to take part in the pattern formation. In this way, activity propagated from the cued location to the end of the contour without spillover to the background.

It might be argued that the need to withdraw the cue signal *J* in order to start the tracing process is a problematic feature of the model. However, we emphasize that *J* should be conceived as a top-down signal arriving from the frontal cortex and not as a component of the feedforward input *I*. In this way, the model is capable of distinguishing between spatial and object-based attention. Sometimes we need to focus on a particular location and other times on all locations occupied by the object. For example, when the target location is known in advance, there is no need to engage object-based attention. In contrast, uncertainty regarding the future target produces object-based effects (Shomstein & Yantis, 2002; 2004; Drummond & Shomstein, 2010). In other words, attention-related activity spreading is not an obligatory process and its engagement depends on the task demands.

The model achieves this distinction by applying *J* to the F-WTA network in a sustained or transient fashion. When *J* is sustained on a location, this location is selected alone. This means that spatial attention holds on to that location. On the other hand, when we solve the problem of whether two locations are connected, we need to focus attention on one location first and then move attention to the second location. In this case, spatial attention transiently visits the first location. We modeled this by transiently switching *J* on and off at a location that is externally designated (by the experimenter) as a starting point for tracing. After *J* is switched off, object-based attention begins to incrementally visit all locations connected with the starting point.

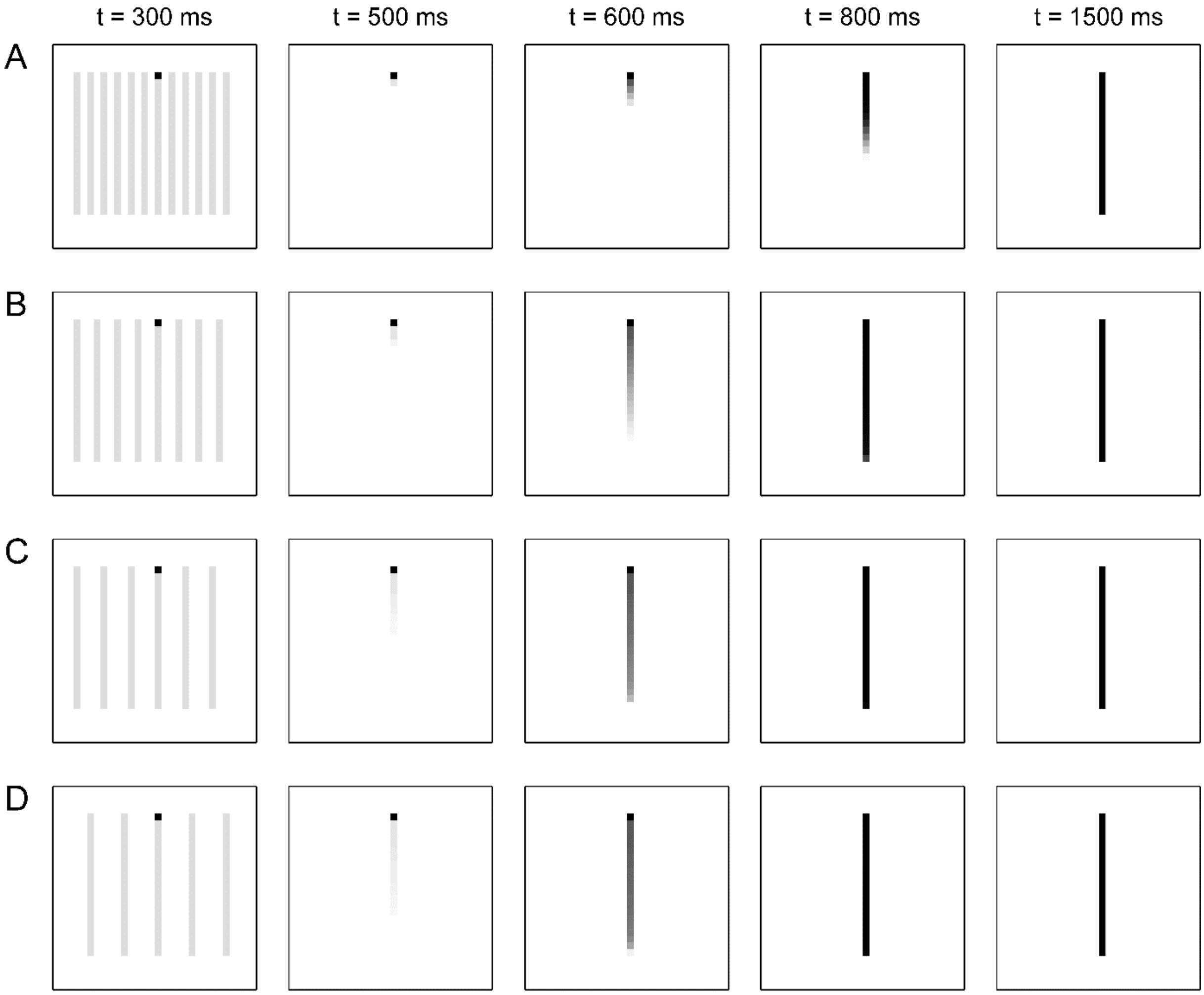


Figure 9. *Simulation of the effect of distractor proximity on the dynamics of the vertical F-WTA network. Each row depicts the F-WTA response to the stimulus shown in Figure 8 with the same corresponding label A–D. Columns depict five representative time points. The spatial cue was applied between t = 200 ms and t = 400 ms. After the cue was withdrawn, at about t = 500 ms, the tracing started with variable speed, depending on the distractor proximity. The tracing was slow with a minute (one-pixel) distance (A) because activity enhancement in the F-WTA network reached the end of the target contour only slightly before t = 1500 ms. The speed of tracing increased with a two-pixel distance (B) where activity enhancement reached the end of the target contour around t = 800 ms. Speed of tracing increased even further with three-pixel (C) and four-pixel distances (D) where activity enhancement reached the end of the target contour slightly after t = 600 ms.*

When we compare the four stimulus conditions in Figure 9, we observe that the F-WTA network automatically adjusted the speed of tracing depending on the proximity of distractor contours. This can be appreciated by observing when enhanced activity reached the bottom end of the target contour in each row. When the separation between contours was only one-pixel wide (Figure 9A), the activity spreading was slow because the F-WTA is allowed to sample signals over the *a_1*-units or the nearest neighbors only. There was no suprathreshold activity at the *c_2*-, *c_3*-, and *c_4*-units, and the larger integration zones offered by the *a_2*-, *a_3*-, and *a_4*-units were disabled. Activity spreading consequently proceeded slowly by advancing one pixel at a time (Figure 9A). It reached the end of the target contour after *t* = 1000 ms but before *t* = 1500 ms. Next, when the separation between contours was two pixels (Figure 9B), the speed of activity spreading doubled because the network advanced along the contour by taking larger two-pixel steps on the target contour. Now, the *c_2*- and *a_2*-units were activated along with the *c_1*- and *a_1*-units, and they allow for a widening of the integration zone of the F-WTA nodes. Consequently, the activity enhancement reached the end of the target contour at about *t* = 800 ms. An even greater speed was achieved when contours were separated by a three-pixel gap (Figure 9C). In this condition, *c_3*-units enabled a wider three-pixel sampling of activity in the F-WTA network. As a result, activity enhancement reached the end of the target contour slightly after *t* = 600 ms. Finally, when the separation between contours was four pixels wide (Figure 9D), the speed of activity spreading was also very high because the F-WTA network received input from the *a_4*-units as a result of the suprathreshold activity of the *c_4*-units. In this condition, the network also reached the end of the target contour slightly after *t* = 600 ms.

Figure 9 reveals another important property of the F-WTA network. As tracing proceeded to the end of the contour, new nodes were recruited and appended to the existing representation of the contour. At the same time, old nodes that were already activated at the earlier stages of tracing remained active to the end of tracing. In this way, the model implements object-based attention that spreads along the whole object, consistent with behavioral findings suggesting that tracing spreads rather than moves along the contour (Houtkamp et al., 2003; Roelfsema et al., 2010; Scholte et al., 2001). Importantly, adding new nodes to the representation of the target curve does not reduce the activity of already activated nodes as would be expected in the WTA network with unconstrained lateral inhibition. This does not take place in the F-WTA network, because of the retrograde signaling, which prevents lateral inhibition among winning nodes irrespective of their total number.

Figure 10 further illustrates the temporal dynamics of the activity of individual nodes in the F-WTA network. Here, we depict the activation of the cued node and the nodes positioned on the target contour that are four, eight, 12, and 16 pixels away from the cued node. Figures 10A–D illustrate the same dynamics as Figures 9A–D, respectively. Again, the same features of the F-WTA dynamics are clearly visible. When the cue was on, the cued location was the only winner. After the cue was removed, the tracing proceeded by the sequential excitation of connected nodes. Each excitatory node attained the same activity level because of the dendritic saturation, which prevented unbounded activity growth. Moreover, presynaptic inhibition mediated by the retrograde signaling prevented lateral inhibition among the winning nodes. Therefore, their activation does not deteriorate as new nodes were recruited into the representation of the same contour. Finally, the speed of tracing depended on the output of the *c*-units. It was slowest when the input to the F-WTA nodes arrives from the c_1-units only (Figure 10A). The tracing speed became increasingly faster when the c_2- and c_3-units were also suprathreshold (Figure 10B and 10C), and it was fastest when all four scales of the CDN were simultaneously active (Figure 10D).

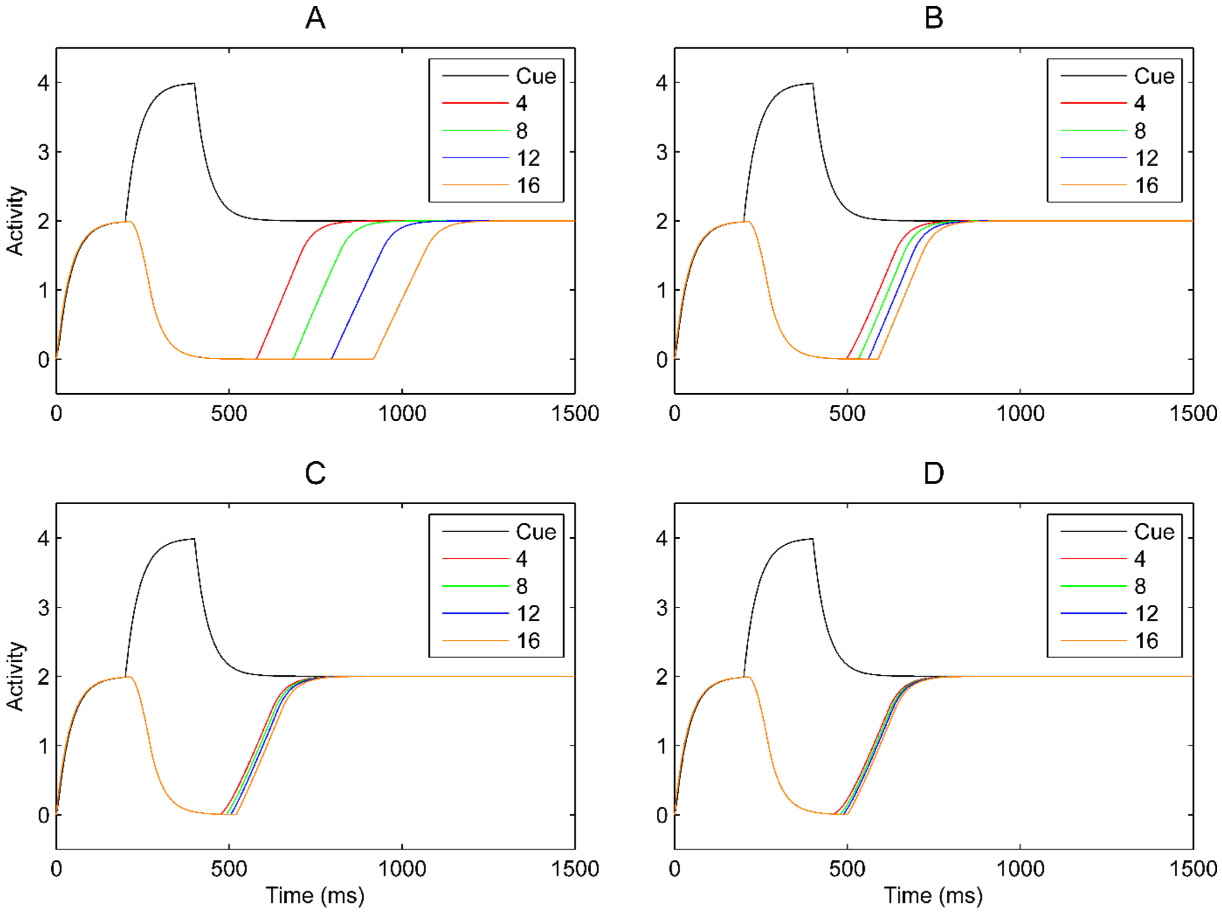


Figure 10. *The temporal dynamics of the activity of the vertical F-WTA nodes lying on the target contour with a target-distractor distance of one pixel (A), two pixels (B), three pixels (C), or four pixels (D). Nodes are labeled by their distances (in pixels) from the cued location.*

To summarize the dynamics of the F-WTA network and to facilitate a comparison with the behavioral data, we calculated the time from the application of the spatial cue to the activation of the set of nodes along the target contour. Also, we added a constant 200 ms to all conditions to take into account processes not related to tracing (e.g., decision making and response preparation). These processes occur after the contour tracing is completed. However, to make Figures 10A–D consistent with this calculation, we inserted the same 200-ms temporal interval between the start of the simulation and the presentation of the cue. In this way, Figures 10A–D depict the total time needed to make response.

Figure 11A displays the time (in milliseconds) that the enhanced activity in the F-WTA network needs to reach the chosen location along the target contour as a function of the distance (in pixels) from the cued location and the proximity between the target and distractor contours. For comparison purposes, Figure 11B displays empirical data from Experiment 3 of Jolicoeur et al. (1991). Two empirically observed trends are also clearly visible in the simulation. The tracing time increases monotonically with the distance from the cue to the target location. Moreover, the slope of the tracing time curve increases as a function of increased proximity between the target and distractors. Interestingly, no appreciable difference exists between tracing times in conditions C and D corresponding to larger target-distractor separation (Figure 11A). The reason is that we employed relatively small input patterns, and the effects of tracing may not become evident until there is a relatively greater distance between the targets along the contour. A similar pattern is also observed in the empirical data (Figure 11B).

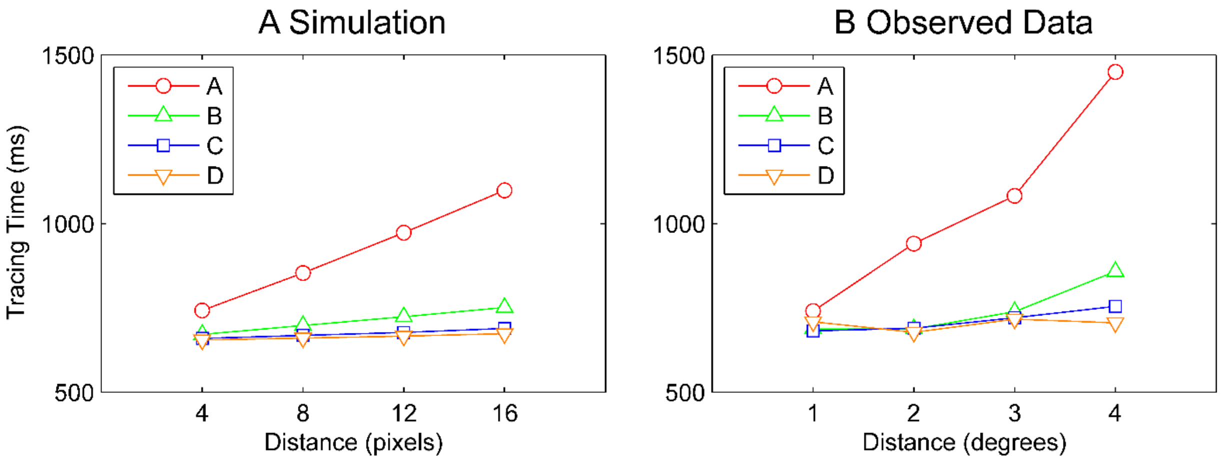


Figure 11. *Comparison between the results of simulation and behavioral data. Simulated tracing times are plotted as a function of distance from the spatial cue (A). A linear trend is observed with a varying slope for different target-distractor distances. To compare the simulation with empirical data, in (B), we redraw mean tracing times for the same response observed in Experiment 3 of Jolicoeur et al. (1991). In their data, distance is measured in degrees of visual angle.*

3.2. The Effect of Contour Curvature

Jolicoeur et al. (1991) also found that at a fixed distractor proximity, tracing becomes increasingly slower as the curvature of the contour increases. With low-resolution input images used here, we emulated the contour curvature by small horizontal displacements of the segments of the vertical contour. Therefore, the contour curvature was defined as the number of horizontal displacements within the same contour size. Figure 12 illustrates four different contour curvatures (including the straight contour) and the corresponding output of the *c*-units in the CDN. In a stimulus with high-curvature contours (Figure 12A), contour segments were displaced horizontally four times to the right and then four times to the left (if we traverse through the grating from top to bottom). Displacements were always by one pixel. Such displacements produce short contour segments that can activate only c_1-units. On the other hand, the c_2-, c_3-, and c_4-units were inactive because of the scale-dependent thresholding. With the medium-curvature contours (Figure 12B), the number of displacements was decreased. There were three displacements to the right and three to the left. This manipulation creates longer contour segments that can activate c_2-units along with the c_1-units. In the stimulus with low-curvature contours (Figure 12C), there was only one displacement to the right and one to the left. Such displacements leave the contour segments large enough to activate c_3-units along with the c_1- and c_2-units. Finally, an input image consisting of straight contours activates all four scales of the CDN (Figure 12D).

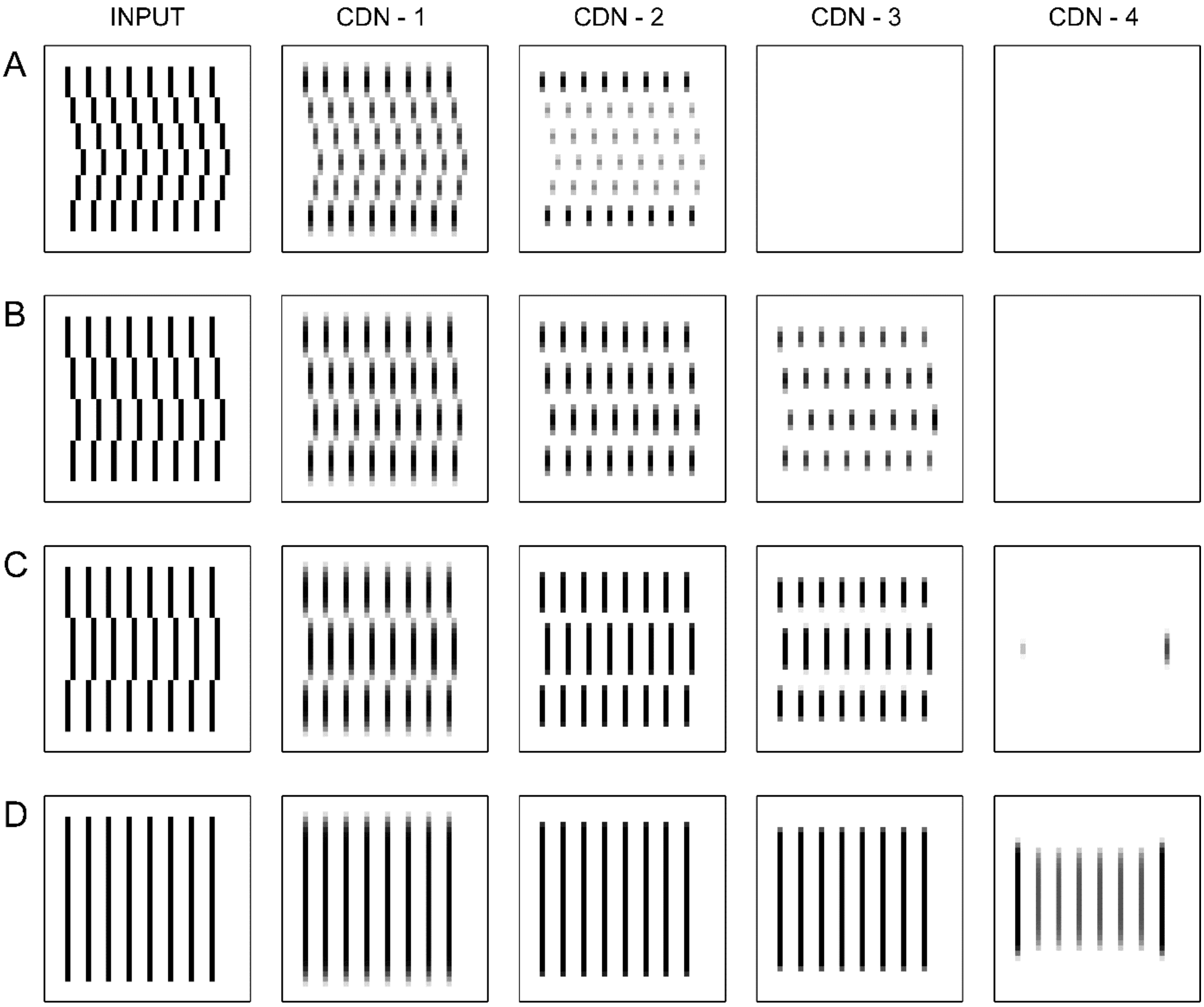


Figure 12. *Simulation of the effect of contour curvature on the output of the vertical c-units at all spatial scales s = {1, 2, 3, 4}. Input is a vertical grating with large (A), medium (B), and small (C) contour curvatures or a straight line (D).*

Figure 13 displays snapshots of the F-WTA activity taken at five representative time points in response to four input configurations illustrated in Figure 12. Each row in Figure 13 depicts the F-WTA response to one stimulus condition: high-curvature (A), medium-curvature (B), and low-curvature contour (C), and the straight line (D). As in the previous simulation, the contour in the middle of the image was chosen for tracing. The spatial cue was applied to the top of the contour placed in the middle of the input pattern. The cue enabled the network to behave similarly to a standard WTA network and to select only the cued location (*t* = 300 ms, across all rows). However, after the cue was removed, tracing began by selectively amplifying the activity of neighboring units that also received feedforward input (*t* = 500 ms, across all rows).

Importantly, the speed of tracing was modulated by the response of the *c*-units in the same way as in the previous simulation. This is seen in Figure 13 by inspecting each row from left to right and by noting when the spreading of enhanced activity reached the lower end of the target contour. In the high-curvature condition (Figure 13A), only a_1-units contributed to the activity spreading because only the c_1-units generated the suprathreshold activity in the CDN. As a result, enhanced activity reached the end of the target contour after t = 1000 ms and before t = 1500 ms. In the medium-curvature condition (Figure 13B), activity enhancement was faster because it advanced by two pixel steps along the contour provided by the activation of the c_2-units. In this case, the end of the target contour was reached slightly after t = 1000 ms. In the low-curvature condition (Figure 13C), activity enhancement proceeded even faster, that is, by three pixel steps. Here, the c_3-units further increased the step size by which activity spreading advanced along the contour, thereby leading to an even shorter time to reach the end of the target contour between t = 800 ms and t = 1000 ms. Finally, with straight contours (Figure 13D), c_4-units were also activated. In this case, activity spreading was fastest and reached the end of the target contour at around t = 800 ms.

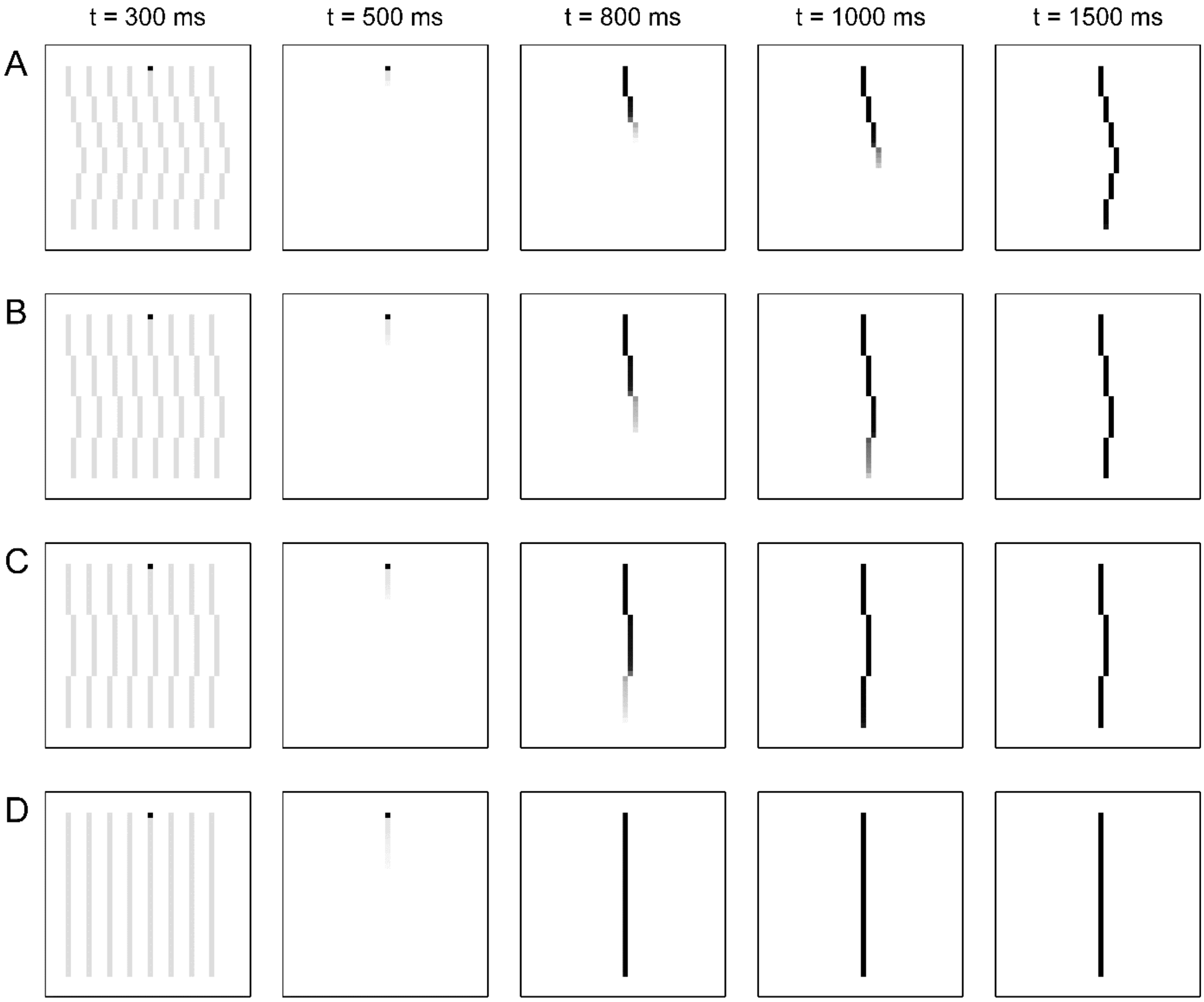


Figure 13. *Simulation of the effect of contour curvature on the dynamics of the vertical F-WTA network. Each row depicts the F-WTA response to the stimulus shown in Figure 12 with the same corresponding label A–D. Columns depict five representative time points. The spatial cue was applied between t = 200 ms and t = 400 ms. After the cue was withdrawn, at about t = 500 ms, the tracing started with variable speed, depending on the contour curvature. Looking from left to right in each row, it is clear that activity enhancement moved slowly on the target contour with large curvature (A) where it reached the end of the contour just before t = 1500 ms. Activity enhancement became increasingly faster with medium (B) and small curvatures (C), where it reached the end of the target contour between t = 800 and t = 1000 ms. Activity enhancement was fastest on the straight line (D) where it reached the end of the target contour before t = 800 ms.*

Figure 14 illustrates the temporal dynamics of the activity of individual nodes in the F-WTA network in the same manner as Figure 10 in Section 3.1. In particular, Figure 14 displays the activation of the cued node and the nodes positioned on the target contour that are 8, 16, and 24 pixels away from the cued node. Labels A, B, C, and D correspond to the high-curvature, medium-curvature, low-curvature, and straight contour condition, respectively.

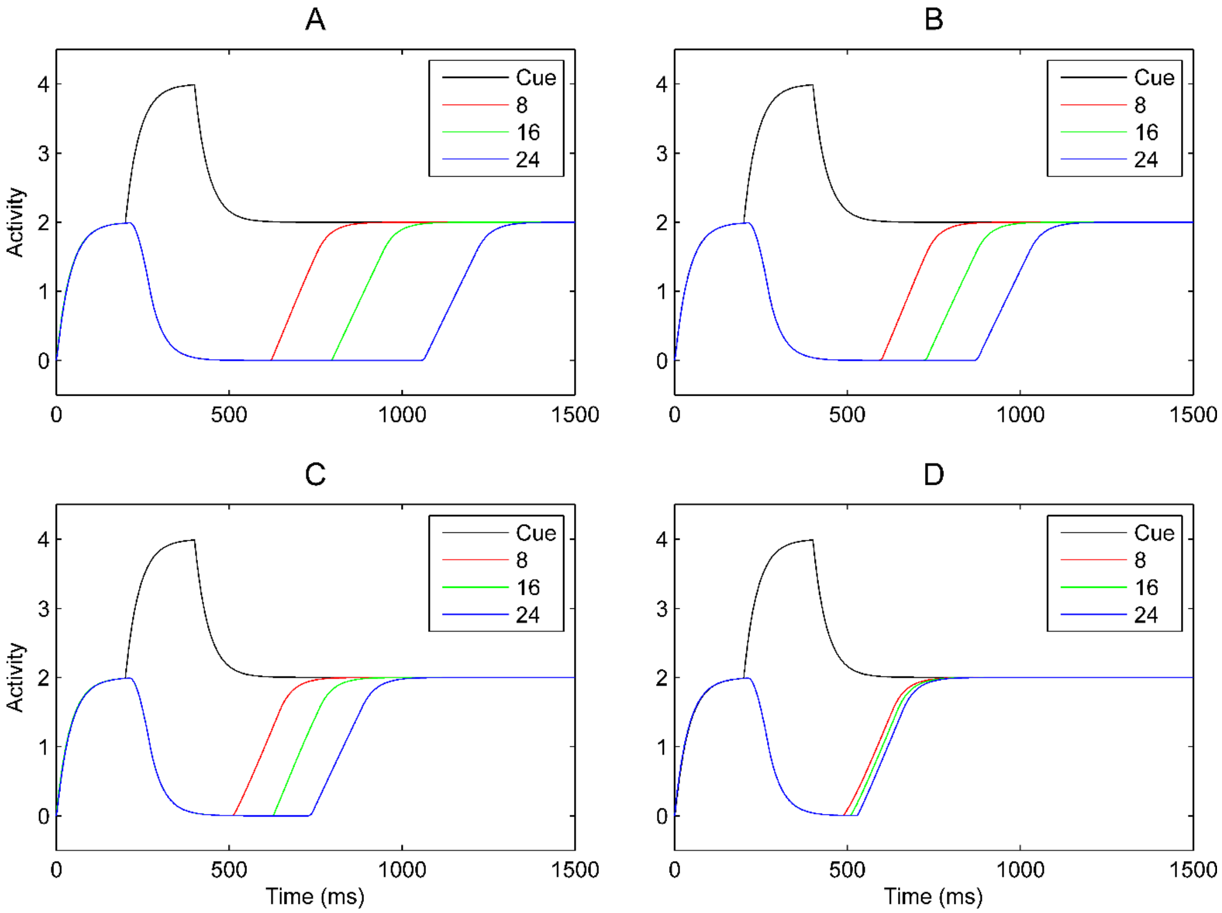


Figure 14. *The temporal dynamics of the activity of the vertical F-WTA nodes lying on the target contour with large (A), medium (B), and small (C) contour curvatures, and the straight line (D). Nodes are labeled by their Hamming distance (in pixels) from the cued location.*

To summarize the results of the simulation, we computed tracing times for selected distances from the cue in the same way as in the previous simulation. Figure 15A demonstrates that contour curvature modulates the slope of the tracing time curves when they are drawn as a function of the cue-target distance. With higher contour curvature, contour tracing becomes increasingly slower. This is consistent with the behavioral data observed in Experiment 1 of Jolicoeur et al. (1991); that data is redrawn in Figure 15B for comparison. Importantly, the spacing between contours is kept constant, and it cannot contribute to this effect.

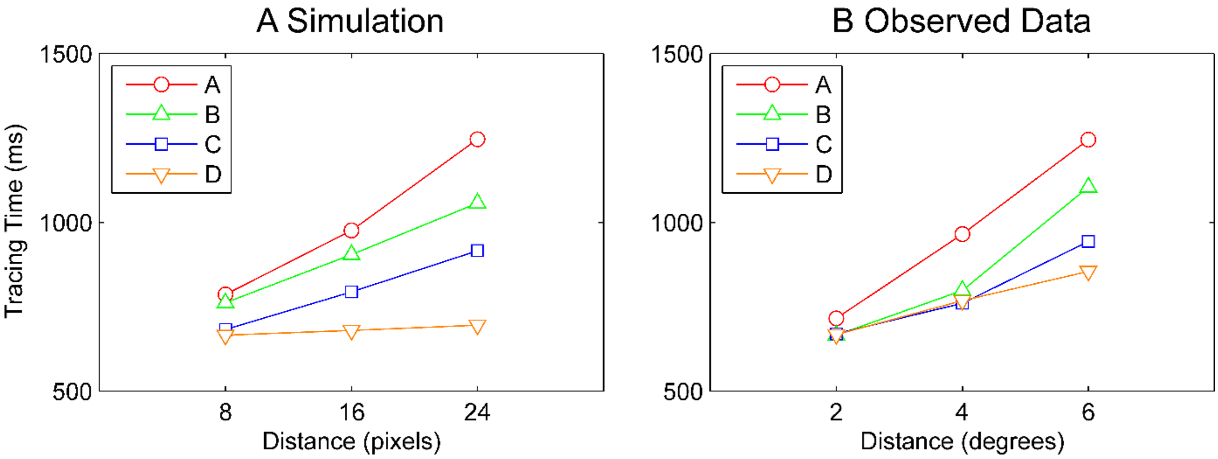


Figure 15. *Comparison between the results of simulation and behavioral data. The simulated tracing times are plotted as a function of distance from the spatial cue (A). The approximately linear trend is observed with a varying slope for different levels of curvature. To compare the simulation with empirical data, in (B), we redraw mean tracing times for the same response observed in Experiment 1 of Jolicoeur et al. (1991). In their data, distance is measured in degrees of visual angle.*

It is interesting to note that the speed of tracing was generally slower in the current simulation relative to the previous one when the same parameter set was used. The reason for this slowing down is the fact that *c*-units are not maximally activated by the curved contours. In other words, segments of the curved contour are not perfectly aligned with the excitatory part of the Gabor filter. Therefore, the F-WTA nodes received smaller total activation from the *c*-units, leading to a slower tracing speed. On the other hand, the data of Jolicoeur et al. (1991) reveals comparable tracing times along curved and straight contours. To account for this discrepancy, we set the synaptic weight of the smallest scale ω_1 to a higher value in the current simulation relative to the previous one. The justification for this modification is the fact that curved contours would likely activate several similar orientations within the more realistic version of the CDN with multiple orientations. Another possibility is that there exist dedicated curve or angular detectors, as proposed by Craft et al. (2007; see also Zhang & von der Heydt, 2010), whose activation may boost the speed of tracing along curved contours by providing an additional source of activation to the F-WTA network. These considerations led to the choice of $\omega_1 = 1$ in simulations involving straight vertical contours (simulations 3.1., 3.3., and 3.4.) and $\omega_1 = 2$ in all simulations involving curved contours (simulations 3.2., 3.5., 3.6., 3.7., and 3.8.).

In simulations 3.1. and 3.2., the spatial cue was always positioned at the beginning of the contour. This does not have to be the case. For example, it is possible that the spatial cue is applied in the middle of the contour or at any other location along the contour. In a similar vein, Crundall, Cole, et al. (2008) suggested that observers sometimes skip the tracing over an irrelevant part of the contour. In other words, they move their spatial attention along the contour and engage it at the location that is closer to the end of the contour. In this way, the tracing is substantially speeded up. The F-WTA network can support such a strategy because it can be cued anywhere on the target contour. After the withdrawal of the cue, the network will start to spread enhanced activity along the cued contour in both directions, namely up and down (for vertical contours) or left and right (for horizontal contours).

3.3. The Effect of Object-based Attentional Cueing

The neural network for attentional selection should be sensitive to abrupt changes in the input because they might be informative and beneficial in solving cognitive tasks, such as examining whether two dots lie on the same contour. McCormick and Jolicoeur (1992) investigated the relationship between attentional cueing and mental contour tracing. They employed grating stimuli similar to those used in the Jolicoeur et al. (1991) study and illustrated in Figure 2A. Object-based cueing involves a temporary brightening of one of the contours in a grating. Importantly, the cue disappears before the dots, whose connectedness should be established, appear on the grating. In particular, the cue remained on the screen for 50 ms, followed by a 67-ms interval of presentation of the original grating without the cue. After the interstimulus interval, the dots appeared on the contour for another 150 ms.

McCormick and Jolicoeur (1992) found that an object-based cue substantially reduces the speed of tracing. In fact, the tracing times to reach different contour segments were almost a flat function of the distance from the start of the contour. Such results suggest that tracing was not necessary and that all segments of the cued contour were selected together in one processing step. Furthermore, it implies that attention remained on the cued contour even after cue disappearance, possibly because of the involvement of spatial working memory.

The simulation presented in parallel in Figures 16 and 17 illustrates that the F-WTA network can handle object-based cueing. Figure 16 displays the activation of the vertical F-WTA network in response to the object-based cue. Another perspective on the same network dynamics is depicted by Figure 17, which reveals the temporal evolution of the activity of a single node at the target and at the distractor contour. In both figures, the same letter denotes the same stimulus condition. Figures 16A and 17A indicate that when the input amplitude of all locations occupied by the target contour is temporarily increased from I_{ij} = 1 to I_{ij} = 3, for the interval of 200 ms, the network selected a cued contour as the winner. In contrast to other WTA networks, which select only a single location (Itti & Koch, 2001), the F-WTA selects all cued locations together. Importantly, after the cue was withdrawn, the cued contour remained selected, illustrating the network's ability to hold the target representation active even after the cue disappeared. In this way, the network is capable of integrating signals that appear on the same location at different times. Further examples of this behavior are presented in Marić and Domijan (2018).

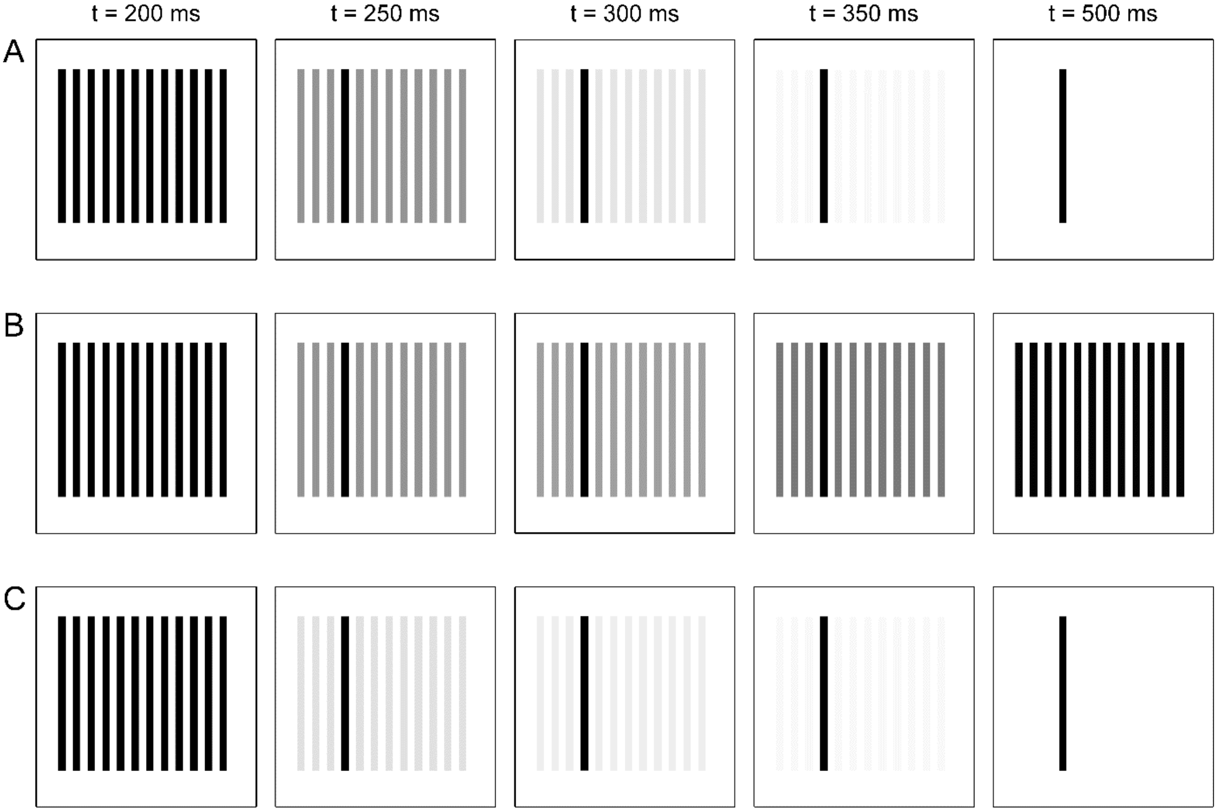


Figure 16. *Simulation of the effect of object-based spatial cueing on contour tracing (McCormick & Jolicoeur, 1992). When the cue duration is 200 ms, the F-WTA network successfully selects the cued contour and stored it into working memory (A). If the cue is presented too quickly (50 ms), the F-WTA network fails to segregate the target contour from the distractors and returns to the state where all contours are selected together (B). However, when the cue strength increases, the F-WTA network is again capable of segregating the target form distractors even under a short (50 ms) cue duration (C).*

We also examined temporal constraints on object-based cueing because McCormick and Jolicoeur (1992) used a much shorter cue duration. Figures 16B and 17B illustrate what occurred when the cue was presented for only 50 ms. In this case, the F-WTA network failed to segregate the target from distractors, and it returned to the initial state where all contours are selected together. The reason was that the activity of the inhibitory interneuron tracked the activity of excitatory nodes with a maximal input magnitude and consequently lagged behind them. When the inhibitory interneuron failed to reach an activity level that is sufficiently above the activity of the excitatory nodes encoding distractor contours, there was no segregation between the target and distractors. One way in which to circumvent this limitation is to apply a much stronger input increment (Figure 16C and 17C). Here, we increased the input amplitude from $I_{ij} = 1$ to $I_{ij} = 7$ for 50 ms on all locations occupied by the target contour. In this case, the network regained its ability to segregate the target from distractors. This additional strength in the input to the F-WTA network may arise from a transient channel that selectively responds to sudden changes in input amplitude (Kulikowski & Tolhurst, 1973; Legge, 1978).

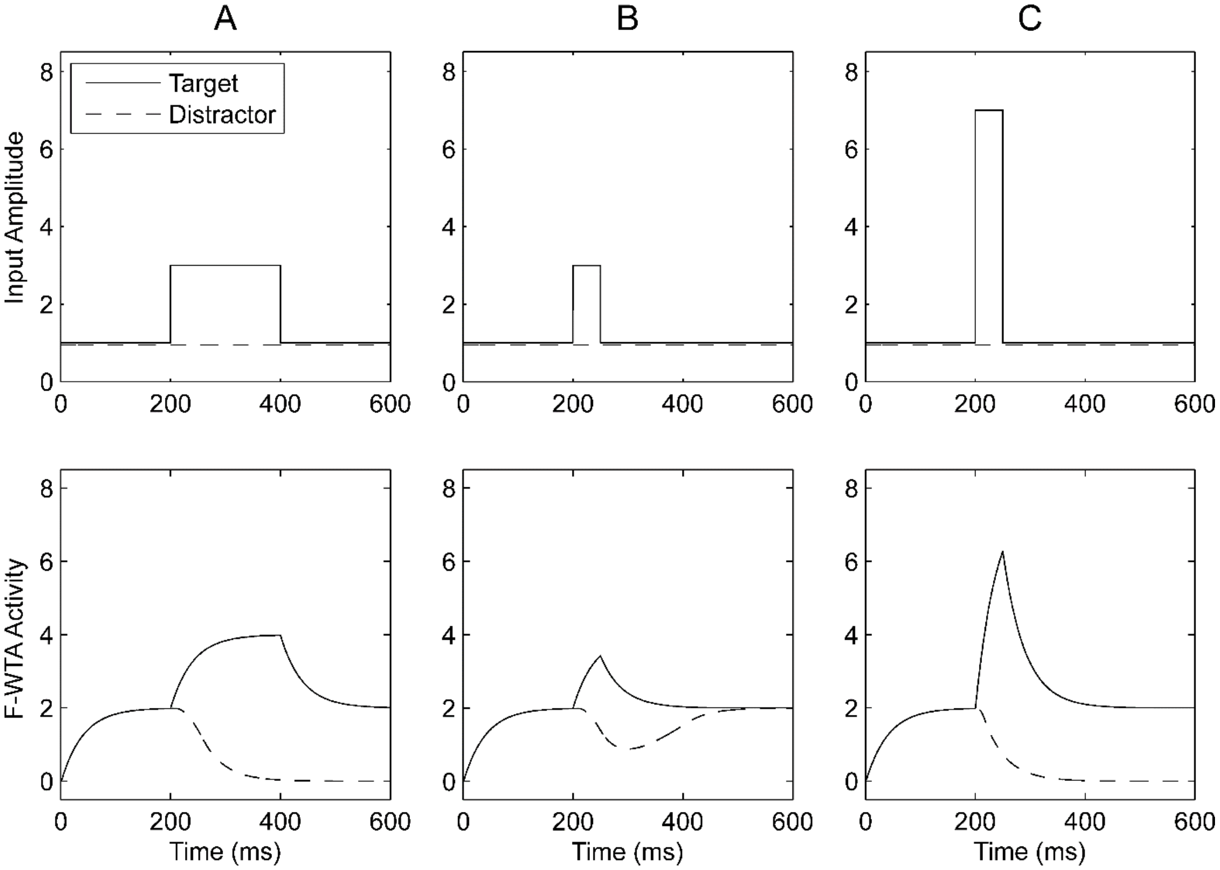


Figure 17. *Another view of the simulation of object-based spatial cueing. Here, the input (top row) and the activity of the F-WTA network (bottom row) at one target and one distractor location are depicted as a function of time. In (A), the F-WTA network successfully segregated the target from the distractor but failed to do so in (B) because of the short presentation of the cue. When the strength of the cue is increased, the F-WTA can segregate the target from the distractor even with a short cue duration (C).*

3.4. Tracing Across Gaps

Ullman (1984) noted that gaps on the contour may disrupt tracing. They may arise from noise in the process of image registration that interferes with the cognitive computations. To illustrate how the F-WTA network handles gaps on the contour, we ran a simulation with the stimuli similar to those used in Section 3.1., but with two small gaps on all contours. The first gap is only one pixel wide, and it was placed on the contours near the starting point of tracing. The second gap is two pixels wide, and it was placed farther away from the cued location on the contours. The dynamics of the F-WTA network is presented in Figure 18.

When the contours were closely spaced (Figure 18A), the F-WTA network got stuck at the first gap. Here, the excitatory lateral interactions were driven by the a_1-units only, since only the c_1-units generated suprathreshold activity. The a_1-units do not have a sufficient lateral extent to cross even the smallest gap. When the contour spacing was slightly larger (Figure 18B), activity spreading can cross the first gap; however, it got stuck at the second gap. In this condition, a_1- and a_2-units were both activated as a result of the activation of their *c*-unit counterparts. The a_2-units have a sufficient lateral extent to cross the first gap but not the second. Finally, when the spacing between contours was wide enough to activate c_3-units also, the corresponding a_3-units helped the activity spreading in the F-WTA network to cross both gaps and to reach the end of the contour (Figure 18C). To cross even larger gaps, the F-WTA network would require the recruitment of the c_4-units or even larger spatial scales (Houtkamp & Roelfsema, 2010; Ullman, 1984). Alternatively, it is possible to enhance the model of the *c*-units with mechanisms for long-range collinear contour detection (Ursino & La Cara, 2004) or with mechanisms for the perception of illusory contours such as the cooperative-competitive loop, which is capable of completing large gaps between collinear contour segments (Gove et al., 1995).

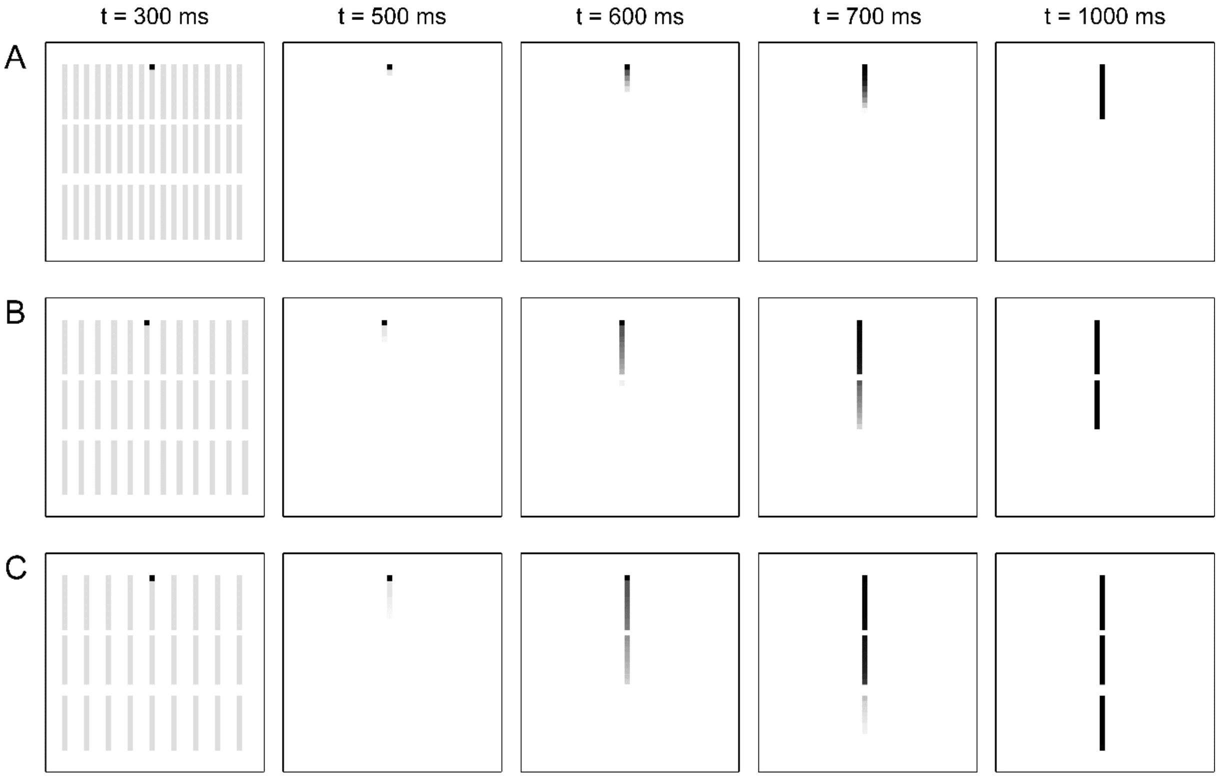


Figure 18. *Simulation of contour tracing across small gaps. The stimuli were similar to those used in the simulation presented in Figure 8. When contours are closely spaced, the activity enhancement in the F-WTA network cannot cross the small gap near the start of the tracing (A). With larger spacing among contours, the activity enhancement can cross the small gap; however, it gets stuck at larger gaps positioned farther along the contour (B). If the spacing between contours is increased even further, then the activity enhancement in the F-WTA network can cross both gaps and complete the spatial representation of the target contour.*

3.5. Solving the Spiral Problem

In previous simulations, we used simple grating stimuli that required vertical boundary detectors only. To demonstrate that the proposed network can also trace more complex patterns composed of vertical and horizontal contour segments, we employed the famous spiral problem (Minsky & Papert, 1988). It was designed to examine the pattern separation capabilities of perceptrons and related learning algorithms. Importantly, the spiral problem has also inspired much of the research on mental contour tracing. Like perceptrons, human observers cannot immediately perceive whether there are two spirals that can be separated or whether there is a single connected spiral. To solve this task, we need to focus on one part of the spiral and spread attention to all connected elements.

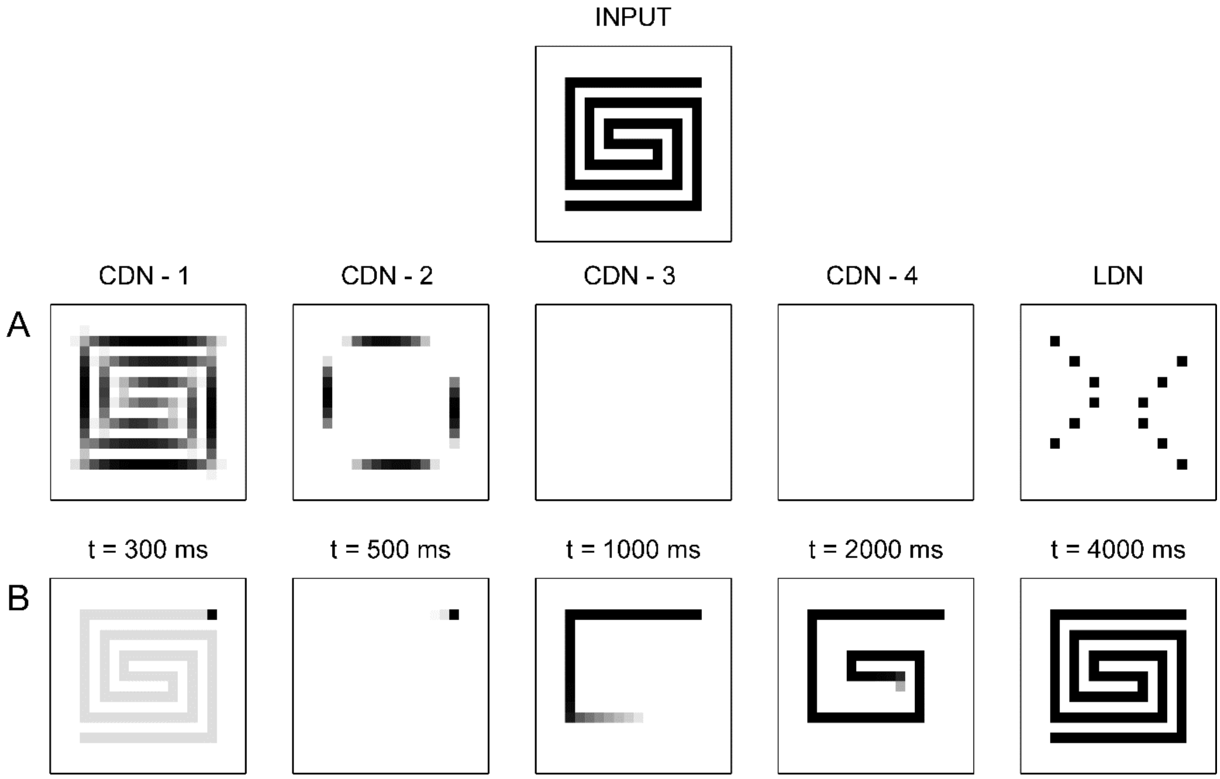


Figure 19. *Simulation of the solution to the spiral problem when the input is a one-spiral image. In (A), the outputs of all four scales of the CDN are presented alongside with the output of the LDN. In (B), snapshots of the F-WTA network activity were taken at five representative time points.*

Figures 19 and 20 illustrate how the F-WTA network solves the spiral problem. The input pattern is one (Figure 19, INPUT) or two spirals (Figure 20, INPUT). To simplify the presentation of the output of the *c*-units and the nodes of the F-WTA network (Figures 19B and 20B), here and in the simulations presented in the next two sections, we used the following convention. We computed the maximum of the outputs of the vertical and horizontal *c*-units at each pixel and displayed it in the image. The same convention was applied in displaying the output of the horizontal and vertical F-WTA networks at selected time steps in Figures 19C and 20C. Also, we included the response of the *l*-units (LDN in Figures 19B and 20B), as their output is important in tracing across L-junctions.

The simulation followed the same scenario as described in Section 3.1. This means that the external spatial cue was applied and then withdrawn. The cue was applied at the top right pixel of the contour, although this is not crucial. After the removal of the cue, the F-WTA network began to trace the contour, as already demonstrated before (Figures 19C and 20C). In Figure 19C, activity enhancement progressively spread along the cued contour from t = 500 ms across *t* = 1000 ms and *t* = 2000 ms until it reached the end of the contour shortly before *t* =

4000 ms. In this case, the steady state activity of the F-WTA network suggests the presence of a single connected spiral.

In Figure 20C, activity enhancement progressed in a similar fashion. However, it reached the end of the cued contour before *t* = 2000 ms because this contour was much shorter relative to the contour used in Figure 19. In Figure 20C, the cued contour was clearly segregated from the distractor at *t* = 2000 ms and afterwards. In this way, the steady state activity of the F-WTA network suggests the presence of two separated spirals.

Importantly, at the corners of spirals, activity spreading turned from the horizontal to the vertical segment, and vice versa. This was a consequence of the activation of the *l*-units (LDN in Figures 19B and 20B), The *l*-units established the excitatory link between two F-WTA networks and enable enhanced activity to jump from one F-WTA network to another. However, such integration of signals from both orientations creates a potential problem when the network traces patterns with X- or T-junctions. This will be addressed in the next two sections.

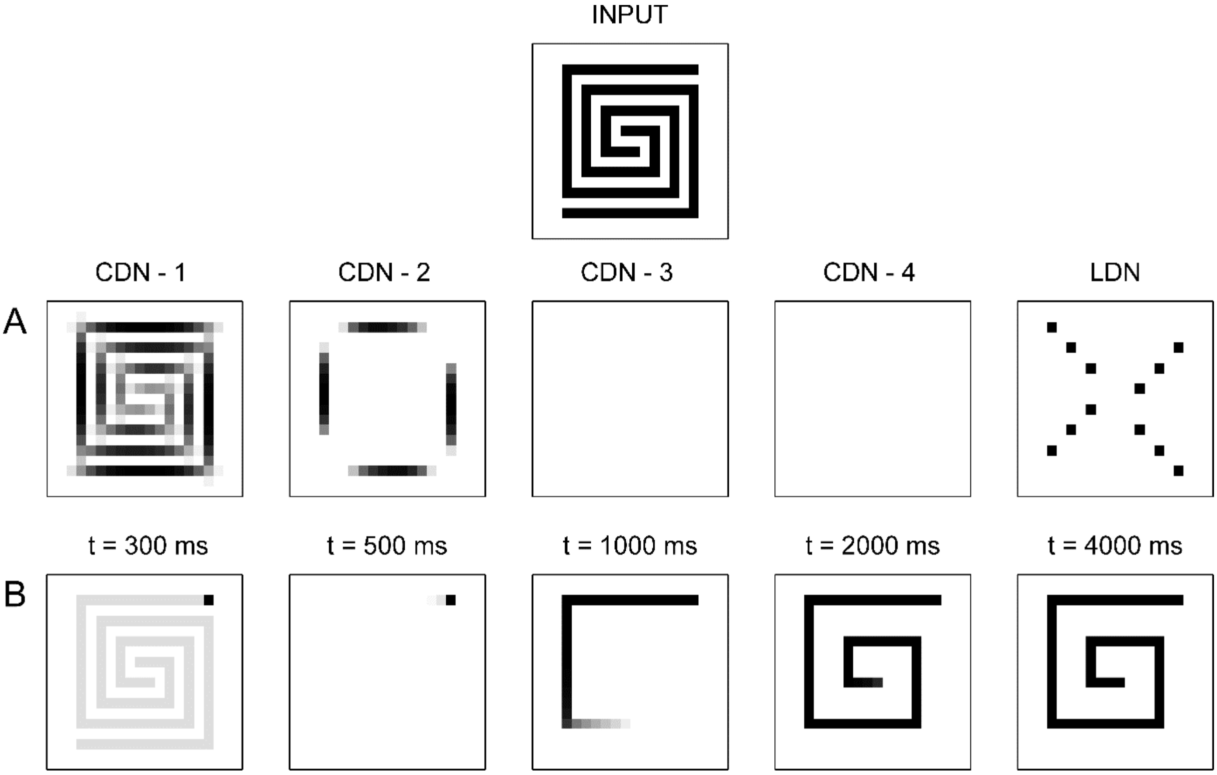


Figure 20. *Simulation of the solution to the spiral problem when the input is a two-spiral image. In (A), the outputs of all four scales of the CDN are presented alongside with the output of the LDN. In (B), snapshots of the F-WTA network activity were taken at five representative time points.*

3.6. Tracing Across X-junctions

Many contour tracing studies employed intersecting contours as stimuli (e.g., Crundall, Cole, et al., 2008; Houtkamp et al., 2003; Scholte et al., 2001). Contour intersections create so-called X-junctions that are especially problematic for contour tracing (Ullman, 1984). When the tracing operator approaches an X-junction from one of the four directions, the direction in which to proceed from the junction onwards is not clear. If there are no orientation-specific mechanisms that constrain tracing, then the tracing would spill over in all directions, leading to a tracing error. Human observers tend to choose the direction that requires the smallest change in orientation. In other words, the tracing will continue in the direction that is collinear to the direction from which it arrives at the X-junction. This is closely related to the well-known Gestalt principle of good continuation (Brooks, 2014). To account for this property, we designed the *l*-unit in a way to respond exclusively to L-junction but not to X-junction although X-junction actually consists of four adjacent L-junctions. However, at the X-junction, there is no end-stopping responses and consequently no *l*-unit activity. As a result, activity in the F-WTA networks continues to propagate in the segregated orientation domains (horizontal or vertical).

Figure 21 demonstrates that the F-WTA networks spread activity enhancement across X-junctions according to the Gestalt principle of good continuation. As input to the network, we used two intersecting squares (Figure 21, INPUT). Figure 21A depicts the responses of CDN and LDN to this input pattern. First, we applied the cue on the upper left square (Figure 21B) using the same cueing procedure as in the previous simulations. After the cue was applied to both horizontal and vertical F-WTA networks, they gradually inhibited all non-cued locations (Figure 21B, t = 300 ms) and eventually selected the cued location alone. After the cue was removed at t = 400 ms (not shown), activity in the horizontal F-WTA network began to spread to the left and right from the cued location (Figure 21B, t = 500 ms). The vertical F-WTA could not exhibit activity spreading at the cued location, as it did not receive input from the vertical *c*-units. However, enhanced activity in the horizontal F-WTA network crossed the L-junctions at the corners of the square and activated the vertical F-WTA network, which continued to propagate enhancement in the vertical direction (Figure 21B, t = 600 ms). Next, activity enhancement passed through the upper right X-junction from above (Figure 21B, t = 700 ms). At the X-junction, there was no activity spillover to the horizontal F-WTA because the *l*-unit was not active here (Figure 21A, LDN). At about t = 800 ms (not shown), the horizontal F-WTA reached another X-junction located in the lower left part of the image and crossed it without spillover. Again, the reason is the absence of the output from the *l*-units. At the end of simulation, both networks reached a steady state with the active representation of the cued square (Figure 21B, *t* = 1000 ms). In the same manner, when the lower right square was cued (Figure 21C, *t* = 300 ms), activity spreading in the F-WTA networks led to the segregation of the cued square from the background without making an error (Figure 21C, *t* = 1000 ms).

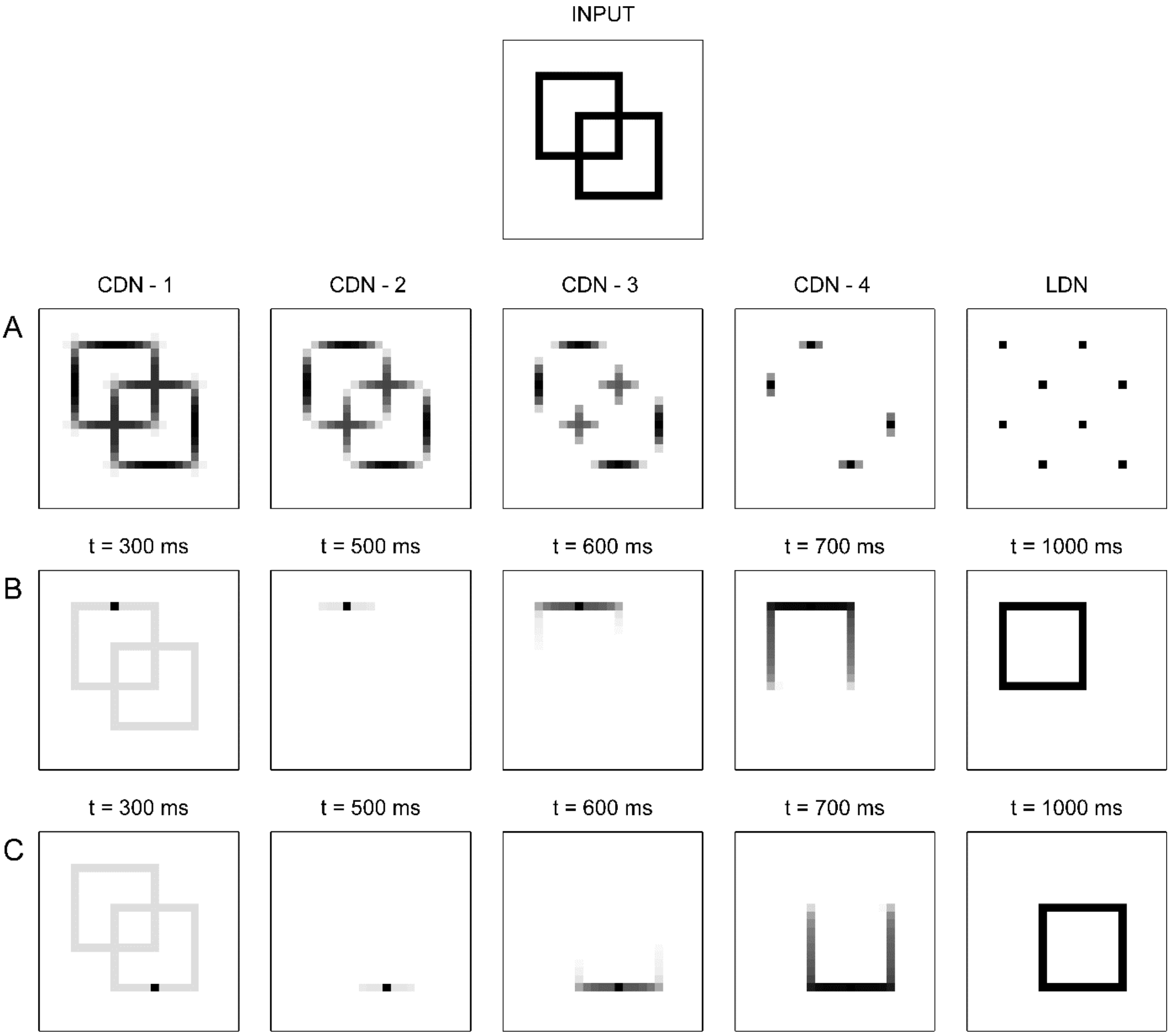


Figure 21. *Simulation of the contour tracing across X-junctions. The input consists of two intersecting squares (INPUT). In (A), the outputs of all four scales of the CDN are presented alongside with the output of the LDN. In (B) and (C), snapshots of the F-WTA network activity were taken at five representative time points. In (B), the spatial cue was applied to the upper square. In (C), the cue was applied to the lower square.*

3.7. Tracing Across T-junctions

Here, we demonstrate that the same effect observed in the previous simulation also occurs at the T-junctions (Figure 22). On the one hand, The T-junctions perceptually signal the presence of an occlusion in depth. In particular, the top of the T is perceived as a part of the object that is closer in depth and that occludes parts of other objects in its immediate neighborhood. On the other hand, the stem of the T belongs to the occluded object (Figure 22, INPUT). The T-junctions are also problematic with respect to tracing because it would be erroneous to join the top and stem of the T-junction together into the representation of the same contour. Here, we demonstrated that the F-WTA networks correctly separate the representation of the occluding or the occluded object depending on the location where the cue is applied. Figure 22A depicts the responses of CDN and LDN to the input pattern.

When the occluding object was cued (Figure 22B), tracing proceeded along its contour without spillover to the contour of the occluded object at the T-junction. In the same manner, when the occluded object was cued (Figure 22C), tracing proceeded along its contour without spillover to the contour of the occluding object. At the T-junction, there was no suprathreshold activity in the LDN and consequently no interaction between the vertical and horizontal F-WTA networks. On the other hand, at the L-junctions, activity spreading crossed between the vertical and horizontal F-WTA networks via a link established by the suprathreshold *l*-units (Figure 22A, LDN). It should be pointed out that the F-WTA network does not have the capacity to represent different depth planes for occluding and occluded objects. Still, it properly detaches them during tracing.

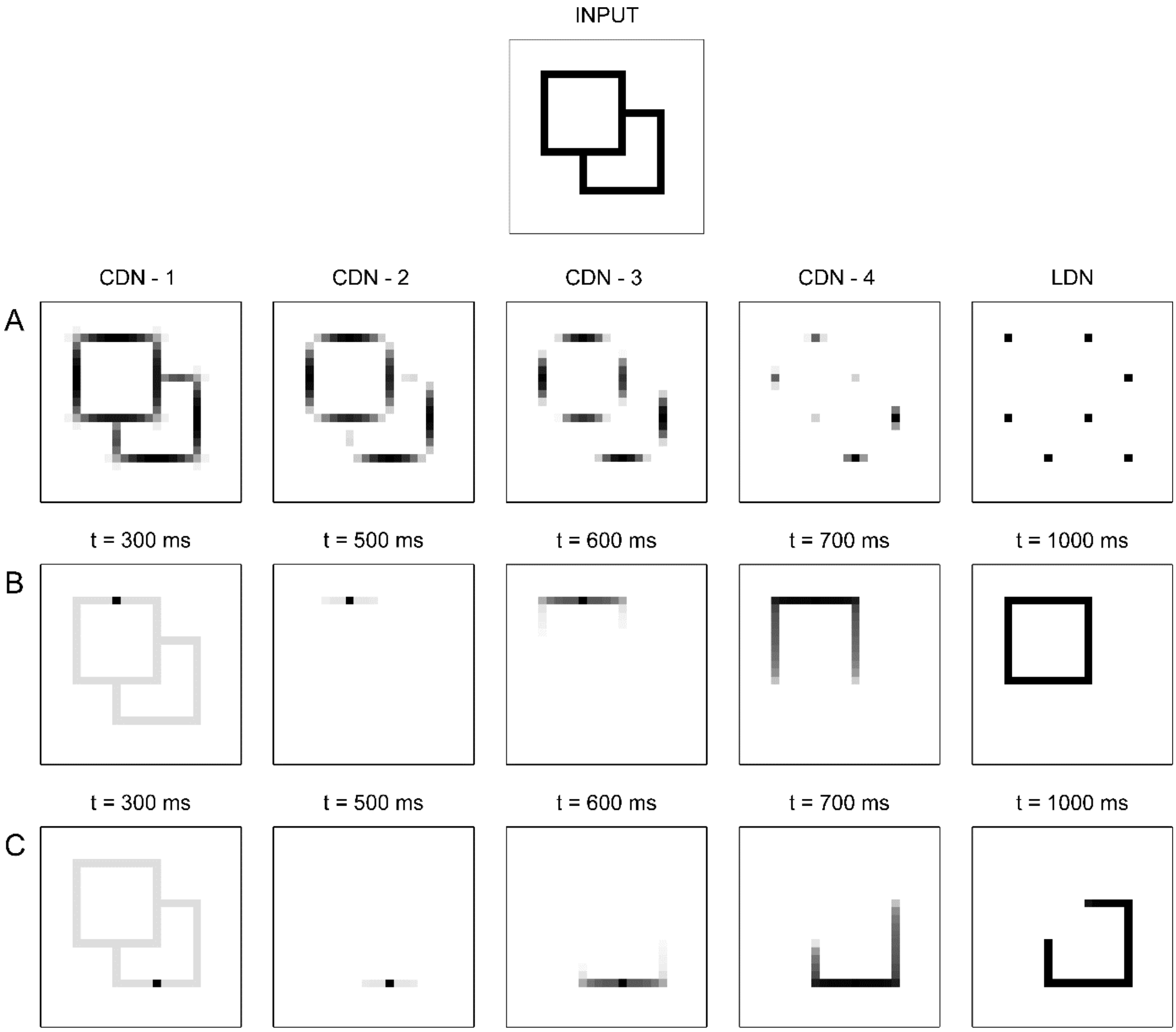


Figure 22. *Simulation of contour tracing across T-junctions. Input consists of two overlapping squares, where one square occludes the other (INPUT). In (A), the outputs of all four scales of the CDN are presented alongside with the output of the LDN. In (B) and (C), snapshots of the F-WTA network activity were taken at five representative time points. In (B), the spatial cue was applied to the occluding square. In (C), the cue was applied to the occluded contour.*

3.8. The Effect of Changes in Parameters

It is important to establish the degree to which the model behavior is affected by changes in the model parameters. To this end, we rerun simulation 3.1. with different parameter settings to examine whether contour tracing could still be observed. Three key parameters are the strength of the dendritic impact on node α, the strength of lateral inhibition β_1, and the strength of the recurrent excitation arriving on inhibitory interneuron β_2. First, we examined systematic changes in one parameter while other parameters were kept fixed at default values.

Parameter α could be set anywhere in the range between 0.5 and 1.5 without affecting the network's ability to perform tracing. In the same manner, weights controlling multiplicative gating from the CDN (ω_1, ω_2, ω_3, ω_4) to the F-WTA network could be set anywhere between 0.2 and 2.0. The weight controlling multiplicative gating from the LDN (ω_L) to the F-WTA network could be set in the range between 1.0 and 4.0. These weights affected the speed of tracing only with higher values, leading to faster tracing. The effect of changes on parameter β_1 was more destructive. Contour tracing was observed when it was set in the range between 0.9 and 1.3. When $\beta_1 < 0.9$, the network settled to a steady state with indiscriminate selection of all nodes that received input. In other words, lateral inhibition was too week and it could not suppress distractors. When $\beta_1 > 1.3$, the network selected the cued location only. In this case, lateral inhibition was too strong, and consequently tracing could not start at all. Parameter β_2 was set at the lower bound to achieve contour tracing, and any value above the default value of 10 led to the same behavior.

Second, we examined joint variations of parameters α and β_1. When β_1 was set at its lower bound of 0.9, α should be in the range between 0.6 and 1.0. When β_1 was in the range between 1.0 and 1.3, α should be set in the range between 0.5 and 1.0. When β_1 was set in the range between 0.9 and 1.3, and if α was set below its prescribed range, the network selected all nodes that received input. On the other hand, if α was set above its prescribed range, the network selected the cued location only. Also, it should be pointed out that thresholds for dendritic and synaptic computation, T_d, T_x, and T_y could be jointly increased or decreased by 50% of their default values without any impact on the observed results. This analysis suggests that the proposed model is quite robust in terms of parameter changes and that it can operate in realistic physiological conditions.

4. DISCUSSION

To account for behavioral findings on contour tracing, McCormick and Jolicoeur (1991, 1994) developed a zoom lens model that captures many of its features. They have identified five component processes that a contour tracing operator should have: (1) a process that can determine whether there is only one contour within the receptive field, (2) a zoom process that can shrink or expand the size of the receptive field until an optimal size is reached such that only one contour remains within it, (3) a process to determine whether the second contour segment has been located, (4) a process that calculates the direction of the next attentional shift, and (5) a process that shifts the receptive field to a new region that contains a small part of the previously selected contour segment. Although successful in explaining patterns of response times, the zoom lens model is purely descriptive. It does not provide details about the possible neurocomputational mechanisms that can support the identified processes. In this regard, we illustrated how interactions between the F-WTA network with two components of the boundary processing, such as CDN and LDN, achieve contour tracing.

The *c*-units detect contour segments of different sizes via multi-scale Gabor filtering. Filters were designed so that the width of their inhibitory part scales with the size of the excitatory part in the preferred direction. Suprathreshold activation of the *c*-units consequently signals that no distractors are present within their respective inhibitory zone. Therefore, the *c*-units implement the process (1). Next, the multiplication between the *a*- and *c*-units gates lateral interactions within the F-WTA network, as illustrated in Figure 5. Therefore, the F-WTA network will make larger or smaller steps along the target contour depending on the activation of the *c*-units of the corresponding size. In this way, multiplicative gating implements process (2), that is, zooming in or zooming out.

Processes from (3) to (5) are intrinsic to the F-WTA network. It automatically finds a path to spread activity along the contour without spillover to empty space around the contour. The reason is that an empty location does not receive sufficient input to cross the QT and to become a part of the representation of the target contour. In contrast, a location occupied by the contour will cross the QT because it receives two independent sources of activation: feedforward input and dendritic output mediating lateral excitation. In other words, its total somatic activation is the same as that of the already activated nodes (see the Appendix for details). In consequence, such a node is incrementally added to the already established representation of the target contour. This process continues until it reaches an empty space. Formally speaking, the network traverses, in its state-space, through a series of forbidden (unstable) subspaces until it reaches a permitted (stable) subspace corresponding to the solution of the problem of connectedness detection (Rutishauser et al., 2015). In addition to the five processes described above, the LDN helps to regulate activity spreading along the L-, X-, and T-junctions, and it prevents spillover to the distractor contour.

The F-WTA network shares a similarity with the filling-in model of perception developed to account for the formation of a surface representation in brightness perception (Grossberg & Todorović, 1988). It was later extended to simulate the detection of connectedness (Grossberg & Wyse, 1991, 1992) and both figure-ground separation and depth perception (Grossberg, 1994). The basic assumption of the filling-in model is that the neural substrate of surface perception is the diffusion of electrical activity among nearest neighbor nodes. Diffusion allows brightness or color signals registered at surface borders to fill in the interior of a surface. However, diffusion is blocked at surface borders because of the separate boundary signals that prevent activity spillover across surfaces.

Two features of the filling-in model are relevant here. First, it is a pixel-by-pixel process that slowly spreads activity enhancement, and it consequently cannot exhibit a variable speed of tracing (Jolicoeur et al., 1991). The exception to this is a model of non-diffusive filling-in proposed by Francis and colleagues (Francis & Ericson, 2004; Francis & Rothmayer, 2003) where lateral interactions extend far in space. However, non-diffusive filling-in does not have the ability to propagate signal enhancement among distant network locations that are outside of the model's receptive fields. Second, both diffusive and non-diffusive models of filling-in have no ability to sustain activity over the target object after input vanishes, because they lack self-recurrent connections. Therefore, they cannot support object-based cueing, as demonstrated by McCormick and Jolicoeur (1992).

The F-WTA network solves these problems by incorporating multiplicative gating of the excitatory interactions and self-recurrent collaterals. Importantly, despite its recurrent connectivity, the F-WTA network remains sensitive to new inputs because feedforward signals arrive on a node's soma, while recurrent interactions between nodes are confined to the dendrite. An additional computational step afforded by the sigmoid nonlinearity in dendrites prevents saturation of the node's activity as it would typically occur in the recurrent excitatory network (Francis et al., 1994). As a consequence, the F-WTA network makes a smooth transition from old to new input. It is even sensitive to the abrupt onset of a new object in the existing input pattern (Marić & Domijan, 2018).

Finally, it should be noted that the F-WTA performs attentional filling-in over contour signals themselves rather than on the empty space enclosed by the contours as it occurs in the filling-in models. Raudies and Neumann (2010) developed a model of figure-ground segregation and object-based attentional selection whereby enhanced neural activity spreads within the interior of object's surface. In the model, attention-related activity enhancement arises from modulating feedback signals arriving from a working memory circuit. Feedback signals traverse through a cortical hierarchy from V4 to V1. In addition, feedback signals spread laterally within each cortical area to encompass the complete surface. Raudies and Neumann (2010) tested the proposed model on various stimulus configurations and made testable predictions regarding the timing of neural responses in model's cortical areas. However, they did not test the effect of target-distractor proximity or target curvature on the speed of tracing, so it is not clear whether their model can handle datasets studied here. Further work is also needed to explore the possibility of adapting the model proposed here to perform attentional filling-in over surfaces instead of contours.

An important limitation of the F-WTA network is that its pattern separation capability is too strong because the distractor contour was suppressed to zero, while the target contour remained active at the level, as it was presented alone. Nevertheless, neurophysiological data suggests only modest modulation; that is, the activity difference between neurons encoding target and distractor contours is small relative to their maximal response (Roelfsema et al., 2003, 1998). However, it should be noted that Roelfsema et al. tested only neurons in V1. On the one hand, there is evidence that the attentional modulation of firing rates is stronger and starts earlier in cortical areas positioned higher in a visual hierarchy such as V4 or inferotemporal cortex (Buffalo et al., 2010; Maunsell & Cook, 2002). Still, neurons in the ventral visual stream respond in a graded fashion, and no evidence was found for all-or-none responses required for WTA selection. On the other hand, the posterior parietal cortex (PPC) is involved in the WTA computation (Bogler et al., 2011). The PPC neurons combine feature (orientation) selectivity with spatial attention (Levichkina et al., 2017; Ogawa & Komatsu, 2009) in order to compute a feature-specific priority map (Veale et al., 2017). Therefore, the PPC may contain a representation of the target contour similar to the output of the F-WTA network.

Another possibility is that the F-WTA network is located in the pulvinar nucleus of the thalamus. The pulvinar is reciprocally connected with the visual areas of the ventral stream, and its deactivation led to a reduction of attentional effects in them (Zhou et al., 2016). Furthermore, the pulvinar is involved in the filtering of distractors because attended objects are encoded with high precision, while there is no measurable encoding of ignored objects (Fischer & Whitney, 2012). Irrespective of its exact anatomical location, we emphasize that the F-WTA network is a part of the attentional network dedicated to target selection and distractor filtering. Such a network may operate in parallel to, and independent of, the network dedicated to visual perception, as indicated by the distinction between A- and N-neurons observed by Roelfsema et al. (2010; see also Beck & Schneider, 2017).

Grossberg and Raizada (2000) and Raizada and Grossberg (2001) demonstrated how response modulation among collinear contour segments arises in a biologically realistic laminar model of perceptual grouping. In their model, activity spreading is the result of the joint activation of horizontal excitatory connections within V1 and V2 and feedback connections from V2 to V1. However, Raizada and Grossberg (2001) also indicated that activity enhancement in the laminar model of perceptual grouping quickly decays as a function of the distance from the start of the tracing. This is not consistent with neurophysiological data showing a comparable amount of response modulation at near and far recording sites relative to the start of tracing (Pooresmaeili & Roelfsema, 2014; Roelfsema et al., 2003) and with behavioral findings showing undiminished capacity to trace spatially extended contours (McCormick & Jolicoeur, 1991, 1994).

Brosch et al. (2015) developed a new learning scheme called RELEARNN (REinforcement LEArning in Recurrent Neural Networks) designed to teach the recurrent network to perform contour tracing. They found that the proposed algorithm requires extensive amount of training (> 100.000 trials) to achieve satisfactory level of performance. In contrast, behavioral studies demonstrated that contour tracing is accomplished without extensive practice. For example, in a study of Jolicoeur et al. (1991), participants completed just a single block of 24 practice trials prior to the start of an experimental session. Moreover, the participants did not receive feedback about their performance during the experimental session. These observations suggest that the participants may engage pre-configured network such as the F-WTA network to solve the contour tracing task. In addition, Brosch et al. (2015) showed that the network trained with the RELEARNN was able to generalize contour tracing to new inputs that had not been presented during training. Interestingly, they also observed that the tracing accuracy dropped from 97% when the network was tested on contours of 5-pixel length to 81% when tested on larger 6-pixel contours. On the other hand, the F-WTA network traces arbitrarily long contours without degrading its neural representation.

An alternative account of mental contour tracing is offered by the selective tuning (ST) model of attention (Rothenstein & Tsotsos, 2014; Tsotsos, 2011; Tsotsos et al., 1995). It was designed to demonstrate how a multi-layered neural architecture, mimicking the hierarchy of visual cortical areas, achieves spatial localization of the most salient object in the input image. Selective tuning operates by making multiple traversals through the visual processing hierarchy. First, task-based priming communicates to the network what to look for and where to find it in the incoming input image. Next, presentation of the stimulus initiates feedforward processing, resulting in a selection of the most salient location. In the next step, feedback processing refines the initial selection and resolves ambiguities about the target location. The ST model was successful in explaining and predicting various behavioral and neural signatures of visuospatial selection (reviewed in Tsotsos, 2011).

Tsotsos and Kruijne (2014) described how to embed ST into a larger neural architecture, called ST-CP, which is capable of implementing complex cognitive programs, including visual routines. On the one hand, in the ST-CP, contour tracing is achieved by discrete jumps of the focus of attention along the contour. To make the jump to the next fixation, the network needs to inhibit representations of already traced parts of the contour. On the other hand, empirical evidence suggests that attention spreads rather than moves along the contour (Houtkamp et al., 2003; Roelfsema et al., 2010; Scholte et al., 2001). This means that representation of the near segment of the contour should remain active while the network traces its far segment (relative to the start), as illustrated in neurophysiological studies (Pooresmaeili & Roelfsema, 2014; Roelfsema et al., 2003). Also, Tsotsos and Kruijne (2014) did not mention how the ST-CP would handle tracing across L-, X-, and T-junctions.

A different approach to the neural basis of perceptual organization is offered by the mechanism of synchronization and de-synchronization of oscillatory activity (Wang, 2005). Neural recordings in the primary visual cortex indicated that synchronization was more likely to occur between cells that are closer in cortical space and between cells that represent similar perceptual features. For example, synchronization of the neural activity in the gamma range (30–100 Hz) is more likely to occur between neurons that have a similar orientation preference, suggesting that such a process is related to perceptual organization and feature binding (Eckhorn, 1999; Singer, 1999; Singer & Gray, 1995). Moreover, synchronization may support the routing of information in the cortex that leads to effective neural communication (Fries, 2005; Salinas & Sejnowski, 2001; Ward, 2003). An example of a computational model implementing principles of neural synchronization is the locally excitatory, globally inhibitory oscillatory network (LEGION) (Wang, 1995; Wang & Terman, 1995, 1997).

Chen and Wang (2001) demonstrated that LEGION can simulate contour tracing when temporal delays are introduced in connections between relaxation oscillators. With delays, the synchronized activity of oscillatory units behaves exactly as the spotlight of attention that moves along the contour. Contrary to this prediction, behavioral and neural data suggest that attention spreads rather than moves along the contour, as already discussed above in the context of the ST-CP model. Furthermore, LEGION can represent the whole contour only when there are no delays. However, without delays, no tracing occurs, and the synchronization of the oscillatory units corresponding to the target contour is achieved implausibly fast. Moreover, contour tracing in LEGION arises from the intrinsic network dynamics and cannot be cued or guided by external sources, as suggested by the results of Crundall, Cole, et al. (2008) and Donovan et al. (2017). In other words, LEGION cannot support state-dependent computation driven by the inputs (Rutishauser & Douglas, 2009).

In a direct test of the involvement of cortical synchrony in contour tracing, Roelfsema et al. (2004) found that synchrony was unrelated to contour grouping because the strength of the synchrony was weaker among neurons encoding the same contour relative to the synchrony among neurons encoding different contours. Instead, they found that the firing rate modulations better correlated with perceptual grouping. More generally, several authors expressed doubts about the possibility that the synchrony of oscillatory activity mediates cognitive computations (Merker, 2013, 2016; Ray & Maunsell, 2015; Shadlen & Movshon, 1999; Thiele & Stoner, 2003).

In summary, we demonstrate that the asymmetric cooperative-competitive network, when embedded in a larger multi-scale architecture for contour and L-junction detection, can support the attentional labeling of connected image elements consistent with the dynamics of mental contour tracing. The model implements scale-dependent attentional filling-in, which incrementally builds a representation of the target contour and simultaneously suppresses representations of the distractor contours. In this way, the proposed model offers a neural interface for the interaction between visual perception and cognition as well as to implement incremental grouping (Roelfsema & Houtkamp, 2011).

ACKNOWLEDGMENT

This work was supported by the University of Rijeka under Grant uniri-drustv-18-177.

APPENDIX. Quenching Threshold in the F-WTA Network

Grossberg (1973) introduced the concept of the QT to describe the property of shunting competitive networks with a sigmoid activation function to enhance the response of all nodes receiving an input amplitude above some critical value and to suppress the activity of all other nodes. In a shunting network, the QT is determined by the network parameters, such as the slope of the sigmoid function. Marić and Domijan (2018) demonstrated that the F-WTA network also has a QT. However, in contrast to the shunting model, the QT in the F-WTA network is dynamic and directly depends on the activity of the excitatory nodes and inhibitory interneuron. To see this, consider the steady state of the F-WTA nodes receiving input with a maximal amplitude denoted by I_{IJ}

$$x_{IJ}^{*} = I_{IJ} + \alpha. \tag{A.1}$$

According to Equations (1) and (2), $x_{IJ}(t) > y(t)$ for all $t > 0$ because $dy/dt = 0$ when $y = x_{IJ} - T_x$. As a consequence, x_{IJ} never receives lateral inhibition from the interneuron because of the strong retrograde inhibition imposed on the pathway from y to x_{IJ}. Furthermore, the recurrent excitatory drive to x_{IJ} is bounded above by the piecewise-linear function given by Equation (4). This is an activation function of the dendrite of the x_{IJ}, which creates an additive gain for the soma of x_{IJ}. Equation (A.1) implies that the set of winning nodes x_{IJ} may be arbitrarily large. However, its size does not affect the capacity of the F-WTA network to faithfully represent all locations receiving I_{IJ}.

Along x_{IJ}, inhibitory interneuron y must also reach its steady state because its activity is driven solely by the inputs from x_{IJ}. The steady state of the inhibitory node y is given by

$$y^{*} = \frac{\beta_2 n\left(x_{IJ}^{*} - T_x\right)}{\beta_2 n + 1}, \tag{A.2}$$

where n is the number of x_{IJ}. When β_2 is chosen to be sufficiently large, and/or there are many nodes with maximal input x_{IJ}, then

$$y^{*} \to x_{IJ}^{*} - T_x\ . \tag{A.3}$$

Node *y* thus computes the maximum over its input. It does not matter how many x_{IJ} exist, because all excitatory drive to *y* vanishes as its activity approaches $x_{IJ} - T_x$.

The x_{IJ} nodes, together with the inhibitory node, create the QT for the rest of the network, which is defined by

$$QT = y^* - T_y = x^*_{IJ} - T_x - T_y \,. \tag{A.4}$$

Equation (A.4) indicates that the QT tracks the activity of the winning nodes. When the QT is set, it divides the remaining excitatory nodes into two subsets depending on whether their activity exceeds QT

$$x^*_{i \neq J, j \neq J} = \begin{cases} I_{ij} + \alpha, & if \quad x_{ij} \geq QT, \\ 0, & if \quad x_{ij} < QT. \end{cases} \tag{A.5}$$

This division is closely related to the behavior of the bistable neurons whose membrane potentials fluctuate between two discrete (so-called UP and DOWN) states (Cossart et al., 2003; Shu et al., 2003). Such neurons may implement AND gate between two independent sources of input signals (Kepecs & Raghavachari, 2007). Here, the QT dictates that both feedforward input and dendrite must be active to enable the node to enter the UP state.

Equations (A.1) and (A.5) demonstrate that the steady state of excitatory nodes tracks the maximal input amplitude. In this way, it is possible to drive the network dynamics externally by modulating the input amplitude. Here, we do that by switching spatial cue J_{ij} on and off. To illustrate the effect of spatial cueing on the QT, we ran a simulation presented in Figure 23. It illustrates the response of a small 1-D version of the F-WTA network with nearest-neighbor connectivity only. The input pattern is as follows

$$I = \begin{bmatrix} 1.0 & .9 & .8 & .7 & .6 \end{bmatrix}. \tag{A.6}$$

Before the cue was applied, all excitatory nodes crossed the QT, even though they received different input amplitudes. According to Equations (A.2) and (A.4), y^* = 1.76 and QT = 1.56. Therefore, $x_i^* > QT$ for all i. After the cue $J_1 = 2$ was applied at t = 300 ms, x_1 activity increased to $I_1 + J_1 + \alpha = 4$. The inhibitory node tracked this activity change and reached a new steady state y^* = 3.45, which raised QT to 3.25. Now, $x_i < QT$ for all $i > 1$, leading to the suppression of all non-cued nodes. After the cue was withdrawn at t = 600 ms, x_1 returned to its former steady state given by $I_1 + \alpha$. Here, we allowed for longer cue duration in order to enable the network to reach its steady state. Withdrawal of the cue lowered the QT to its former value of 1.56. Also, it opened the opportunity for x_2 to cross the QT and to initiate the propagation of activity enhancement or tracing. This simulation also illustrates that, because of its QT, the F-WTA network is capable of tracing across contour segments with variable input amplitudes, as demonstrated by Pooresmaeili et al. (2010).

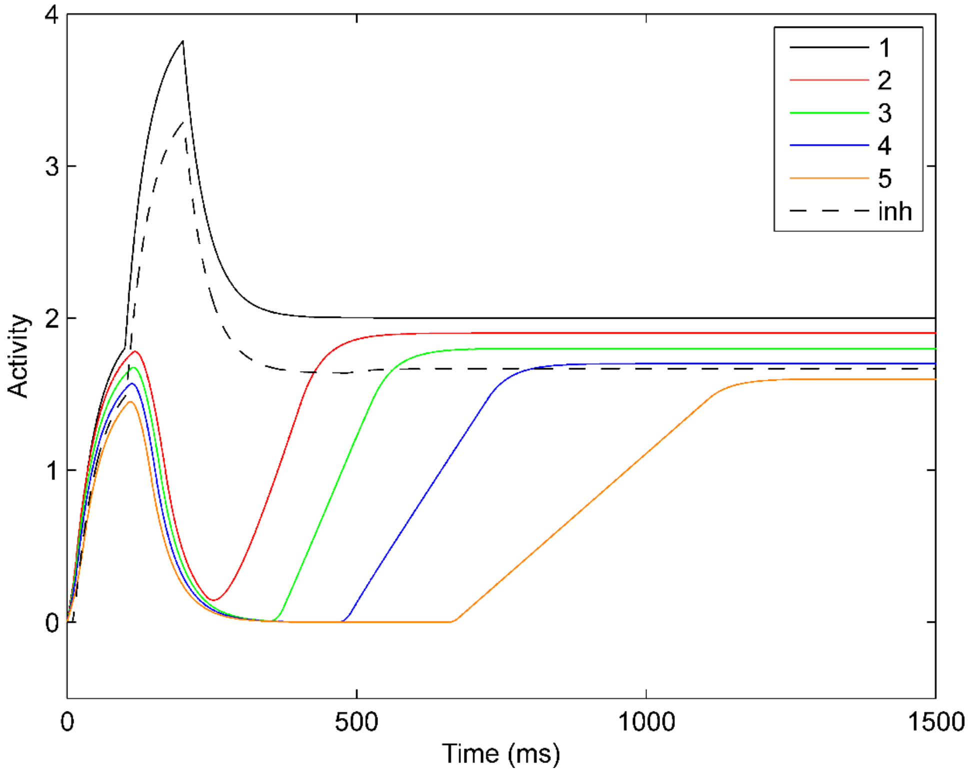


Figure 23. *The relationship between the activity of the excitatory nodes x_1, …, x_5 and inhibitory node y in a response to the input pattern given by Equation (A.6).*

REFERENCES

Abbott, L. F., & Regehr, W. G. (2004). Synaptic computation. *Nature, 431*(7010), 796–803. https://doi.org/10.1038/nature03010

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience, 10*(10), 1313–1321. https://doi.org/10.1038/nn1975

Beck, J., & Schneider, K. (2017). Attention and mental primer. *Mind & Language*, *32*(4), 463–494. https://doi.org/10.1111/mila.12148

Bogler, C., Bode, S., & Haynes, J. D. (2011). Decoding successive computational stages of saliency processing. *Current Biology, 21*(19), 1667–1671. https://doi.org/10.1016/j.cub.2011.08.039

Brooks, J. L. (2014). Traditional and new principles of perceptual grouping. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 57–87). Oxford University Press.

Brosch, T., Neumann, H., & Roelfsema, P. R. (2015). Reinforcement learning of linking and tracing contours in recurrent neural networks. *PLoS Computational Biology, 11*(10), e1004489. https://doi.org/10.1371/journal.pcbi.1004489

Buffalo, E.A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backwards progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America, 107*(1), 361–365. https://doi.org/10.1073/pnas.0907658106

Chen, K., & Wang, D. L. (2001). Perceiving geometric patterns: From spirals to inside-outside relations. *IEEE Transactions on Neural Networks, 12*(5), 1084–1102. https://doi.org/10.1109/72.950138

Cosman, J. D., & Vecera, S. P. (2012). Object-based attention overrides perceptual load to modulate visual distraction. *Journal of Experimental Psychology: Human Perception and Performance, 38*(3), 576–579. https://doi.org/10.1037/a0027406

Cossart, R., Aronov, D., & Yuste, R. (2003). Attractor dynamics of network UP states in the neocortex. *Nature, 423*(6937), 283–288. https://doi.org/10.1038/nature01614

Craft, E., Schuetze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology*, *97*, 4310–4326. https://doi.org/10.1152/jn.00203.2007

Crundall, D., Cole, G. G., & Underwood, G. (2008). Attentional and automatic processes in line tracing: Is tracing obligatory? *Perception & Psychophysics, 70*(3), 422–430. https://doi.org/10.3758/PP.70.3.422

Crundall, D., Dewhurst, R., & Underwood, G. (2008). Does attention move or spread during mental curve tracing? *Perception & Psychophysics, 70*(2), 374-388. https://doi.org/10.3758/pp.70.2.374

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research, 20*(10), 847–856. https://doi.org/10.1016/0042-6989(80)90065-6

Donovan, I., Pratt, J., & Shomstein, S. (2017). Spatial attention is necessary for object-based attention: Evidence from temporal-order judgments. *Attention, Perception, & Psychophysics, 79*(3), 753–764. https://doi.org/10.3758/s13414-016-1265-6

Drummond, L., & Shomstein, S. (2010). Object-based attention: Shifting or uncertainty? *Attention, Perception, & Psychophysics, 72(7),* 1743-1755. https://doi.org/10.3758/APP.72.7.1743

Eckhorn, R. (1999). Neural mechanisms of visual feature binding investigated with microelectrodes and models. *Visual Cognition, 6*, 231-265. https://doi.org/10.1080/135062899394975

Fino, E., Packer, A. M., & Yuste, R. (2013). The logic of inhibitory connectivity in the neocortex. *Neuroscientist, 19*(3), 228-237. https://doi.org/10.1177/1073858412456743

Fischer, J., & Whitney, D. (2012). Attention gates visual coding in the human pulvinar. *Nature Communications, 3*(1051). https://doi.org/10.1038/ncomms2054

Francis, G., & Ericson, J. (2004). Using afterimages to test neural mechanisms for perceptual filling-in. *Neural Networks, 17*(5–6), 737–752. https://doi.org/10.1016/j.neunet.2004.01.007

Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, *34*, 1089–1104. https://doi.org/10.1016/0042-6989(94)90012-4

Francis, G., & Rothmayer, M. (2003). Interactions of afterimages for orientation and color: Experimental data and model simulations. *Perception & Psychophysics, 65*(4), 508–522. https://doi.org/10.3758/bf03194579

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9*(10), 474–480. https://doi.org/10.1016/j.tics.2005.08.011

Fukai, T., & Tanaka, S. (1997). A simple neural network exhibiting selective activation of neuronal ensembles: From winner-take-all to winners-share-all. *Neural Computation, 9*(1), 77–97. https://doi.org/10.1162/neco.1997.9.1.77

Gove, A., Grossberg, S., & Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience, 12*(6), 1027–1052. https://doi.org/10.1017/S0952523800006702

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52*(3), 213–257. https://doi.org/10.1002/sapm1973523213

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics, 55*(1), 48–121. https://doi.org/10.3758/BF03206880

Grossberg, S., Mingolla, E., & Ross, W. D. (1997). Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neuroscience, 20*(3), 106–111. DOI: https://doi.org/10.1016/s0166-2236(96)01002-8

Grossberg, S., & Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research, 40*(10–12), 1413–1432. https://doi.org/10.1016/S0042-6989(99)00229-1

Grossberg, S., & Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception & Psychophysics, 43*(3), 241–277. https://doi.org/10.3758/bf03207869

Grossberg, S., & Wyse, L. (1991). A neural network architecture for figure-ground separation of connected scenic figures. *Neural Networks, 4*, 723-742. https://doi.org/10.1016/0893-6080(91)90053-8

Grossberg, S., & Wyse, L. (1992). Figure-ground separation of connected scenic figures: Boundaries, filling-in, and opponent processing. In G. A. Carpenter & S. Grossberg (Eds.), *Neural Networks for Vision and Image Processing* (pp. 161–194). MIT Press.

Haarmann, H., & Usher, M. (2001). Maintenance of semantic information in capacity-limited item short-term memory. *Psychonomic Bulletin & Review, 8*(3), 568–578. https://doi.org/10.3758/BF03196193

Hansen, T., & Neumann, H. (2004). A simple cell model with dominating opponent inhibition for robust image processing. *Neural Networks, 17*(5–6), 647–662. https://doi.org/10.1016/j.neunet.2004.04.002

Häusser, M., & Mel, B. W. (2003). Dendrites: Bug or feature? *Current Opinion in Neurobiology, 13*(3), 372–383. https://doi.org/10.1016/s0959-4388(03)00075-8

Hollingworth, A., Maxcey-Richard, A. M., & Vecera, S. P. (2012). The spatial distribution of attention within and across objects. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 135–151. https://doi.org/10.1037/a0024463

Houtkamp, R., & Roelfsema, P. R. (2010). Parallel and serial grouping of image elements in visual perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*(6), 1443–1459. https://doi.org/10.1037/a0020248

Houtkamp, R., Spekreijse, H., & Roelfsema, P. R. (2003). A gradual spread of attention. *Perception & Psychophysics, 65*(7), 1136–1144. https://doi.org/10.3758/bf03194840

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114*(3), 599–631. https://doi.org/10.1037/0033-295x.114.3.599

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*, 194. https://doi.org/10.1038/35058500

Jeurissen, D., Self, M. W., & Roelfsema, P. R. (2016). Serial grouping of 2D-image regions with object-based attention in humans. *Elife, 5*. https://doi.org/10.7554/eLife.14320

Jolicoeur, P., & Ingleton, M. (1991). Size invariance in curve tracing. *Memory & Cognition, 19*(1), 21–36. https://doi.org/10.3758/bf03198493

Jolicoeur, P., Ullman, S., & Mackay, M. (1986). Curve tracing: a possible basic operation in the perception of spatial relations. *Memory & Cognition, 14*(2), 129–140. https://doi.org/10.3758/BF03198373

Jolicoeur, P., Ullman, S., & Mackay, M. (1991). Visual curve tracing properties. *Journal of Experimental Psychology: Human Perception and Performance, 17*(4), 997–1022. https://doi.org/10.1037/0096-1523.17.4.997

Kaski, S., & Kohonen, T. (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Networks, 7*(6–7), 973–984. https://doi.org/10.1016/S0893-6080(05)80154-6

Kepecs, A., & Raghavachari, S. (2007). Gating information by two-state membrane potential fluctuations. *Journal of Neurophysiology, 97*(4), 3015–3023. https://doi.org/10.1152/jn.01242.2006

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*(4), 219–227.

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology, 232*, 149–162. https://doi.org/10.1113/jphysiol.1973.sp010261

Legge, G. E. (1978). Sustained and transient mechanisms in human vision: temporal and spatial properties. *Vision Research, 18,* 69–81. https://doi.org/10.1016/0042-6989(78)90079-2

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27,* 889–935. https://doi.org/10.1068/p270889

Levichkina, E., Saalmann, Y. B., & Vidyasagar, T. R. (2017). Coding of spatial attention priorities and object features in the macaque lateral intraparietal cortex. *Physiological Reports, e13136*. https://doi.org/10.14814/phy2.13136

London, M., & Häusser, M. (2005). Dendritic computation. *Annual Review of Neuroscience, 28*(1), 503–532. https://doi.org/10.1146/annurev.neuro.28.061604.135703

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America, 70*(11), 1297–1300. https://doi.org/10.1364/JOSA.70.001297

Marić, M., & Domijan, D. (2018). A neurodynamic model of feature-based spatial selection. *Frontiers in Psychology, 9,* 417 https://doi.org/10.3389/fpsyg.2018.00417

Maunsell, J. H., & Cook, E. P. (2002). The role of attention in visual processing. *Philosophical Transactions of the Royal Society of London. B Biological Sciences, 357*(1424), 1063–1072. https://doi.org/10.1098/rstb.2002.1107

McCormick, P. A., & Jolicoeur, P. (1991). Predicting the shape of distance functions in curve tracing: Evidence for a zoom lens operator. *Memory & Cognition, 19*(5), 469–486. https://doi.org/10.3758/BF03199570

McCormick, P. A., & Jolicoeur, P. (1992). Capturing visual attention and the curve tracing operation. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 72–89. https://doi.org/10.1037/0096-1523.18.1.72

McCormick, P. A., & Jolicoeur, P. (1994). Manipulating the shape of distance effects in visual curve tracing: Further evidence for the zoom lens model. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 48*(1), 1–24. https://doi.org/10.1037/1196-1961.48.1.1

Mel, B. W. (2016). Towards a simplified model of an active dendritic tree. In G. Stuart, N. Spruston, & M. Häusser (Eds.), *Dendrites* (3rd ed., pp. 465–486). Oxford University Press.

Merker, B. H. (2013). Cortical gamma oscillations: The functional key is activation, not cognition. *Neuroscience & Biobehavioral Reviews, 37*(3), 401–417. https://doi.org/10.1016/j.neubiorev.2013.01.013

Merker, B. H. (2016). Cortical gamma oscillations: Details of their genesis preclude a role in cognition. *Frontiers in Computational Neuroscience, 10,* 78. https://doi.org/10.3389/fncom.2016.00078

Minsky, M. L., & Papert, S. A. (1988). *Perceptrons: Expanded edition*. MIT Press.

Ogawa, T., & Komatsu, H. (2009). Condition-dependent and condition-independent target selection in the macaque posterior parietal cortex. *Journal of Neurophysiology*, *101*(2), 721–736. https://doi.org/10.1152/jn.90817.2008

Pooresmaeili, A., Poort, J., Thiele, A., & Roelfsema, P. R. (2010). Separable codes for attention and luminance contrast in the primary visual cortex. *Journal of Neuroscience*, *30*(38), 12701–12711. https://doi.org/10.1523/jneurosci.1388-10.2010

Pooresmaeili, A., & Roelfsema, P. R. (2014). A growth-cone model for the spread of object-based attention during contour grouping. *Current Biology, 24*(24), 2869–2877. https://doi.org/ 10.1016/j.cub.2014.10.007

Pringle, R., & Egeth, H. E. (1988). Mental curve tracing with elementary stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 14*(4), 716–728. https://doi.org/ 10.1037/0096-1523.14.4.716

Raizada, R., & Grossberg, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition, 8*(3–5), 431–466. https://doi.org/10.1080/13506280143000070

Raudies, F., & Neumann, H. (2010). A neural model of the temporal dynamics of figure-ground segregation in motion perception. *Neural Networks, 23,* 160–176. https://doi.org/10.1016/j.neunet.2009.10.005

Ray, S., & Maunsell, J. H. (2015). Do gamma oscillations play a role in cerebral cortex? *Trends in Cognitive Sciences*, *19*(2), 78–85. https://doi.org/10.1016/j.tics.2014.12.002

Regehr, W. G., Carey, M. R., & Best, A. R. (2009). Activity-dependent regulation of synapses by retrograde messengers. *Neuron, 63*(2), 154–170. https://doi.org/b5rzs8

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics, 73*(8), 2542–2572. https://doi.org/d3m47z

Roelfsema, P. R., Houtkamp, R., & Korjoukov, I. (2010). Further evidence for the spread of attention during contour grouping: A reply to Crundall, Dewhurst, and Underwood (2008). *Attention, Perception, & Psychophysics, 72*(3), 849–862. https://doi.org/10.3758/APP.72.3.849

Roelfsema, P. R., Khayat, P. S., & Spekreijse, H. (2003). Subtask sequencing in the primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 100*(9), 5467–5472. https://doi.org/10.1073/pnas.0431051100

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*(6700), 376–381. https://doi.org/10.1038/26475

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience, 7*(9), 982–991. https://doi.org/10.1038/nn1304

Roelfsema, P. R., & de Lange, F. P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Annual Review of Vision Science, 2,* 131–151. https://doi.org/gh4j

Roelfsema, P. R., & Singer, W. (1998). Detecting connectedness. *Cerebral Cortex, 8*(5), 385–396. https://doi.org/10.1093/cercor/8.5.385

Rothenstein, A. L., & Tsotsos, J. K. (2014). Attentional modulation and selection – An integrated approach. *PLoS ONE, 9*(6), e99681. https://doi.org/gh4h

Rutishauser, U., & Douglas, R. J. (2009). State-dependent computation using coupled recurrent networks. *Neural Computation, 21*(2), 478–509. https://doi.org/10.1162/neco.2008.03-08-734

Rutishauser, U., Douglas, R. J., & Slotine, J. J. (2011). Collective stability of networks of winner-take-all circuits. *Neural Computation, 23*(3), 735–773. https://doi.org/d8s62g

Rutishauser, U., Slotine, J. J., & Douglas, R. J. (2012). Competition through selective inhibitory synchrony. *Neural Computation*, 24(8), 2033–2052. https://doi.org/f32k53

Rutishauser, U., Slotine, J. J., & Douglas, R. J. (2015). Computation in dynamically bounded asymmetric systems. *PLoS Computational Biology, 11*(1), e1004039. https://doi.org/10.1371/journal.pcbi.1004039

Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience, 2*(8), 539–550. https://doi.org/d8p39s

Scholte, H. S., Spekreijse, H., & Roelfsema, P. R. (2001). The spatial profile of visual attention in mental curve tracing. *Vision Research, 41,* 2569–2580. https://doi.org/10.1016/S0042-6989(01)00148-1

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron, 24*(1), 67–77, 111–125. https://doi.org/10.1016/s0896-6273(00)80822-3

Shomstein, S., & Yantis, S. (2002). Object-based attention: Sensory modulation or priority setting? *Perception & Psychophysics*, *64*, 41–51. https://doi.org/10.3758/BF03194556

Shomstein, S., & Yantis, S. (2004). Configural and contextual prioritization in object-based attention. *Psychonomic Bulletin & Review, 11*, 247–253. https://doi.org/bqj358

Shu, Y., Hasenstaub, A., & McCormick, D. A. (2003). Turning on and off recurrent balanced cortical activity. *Nature, 423*(6937), 288–293. https://doi.org/10.1038/nature01616

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron, 24*, 49–65. https://doi.org/10.1016/s0896-6273(00)80821-1

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18*, 555–586. https://doi.org/cgx8jp

Thiele, A., & Stoner, G. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature, 421*(6921), 366–370. https://doi.org/10.1038/nature01285

Thielscher, A., & Neumann, H. (2008). Globally consistent depth sorting of overlapping 2D surfaces in a model using local recurrent interactions. *Biological Cybernetics*, 98, 305–337. https://doi.org/10.1007/s00422-008-0211-7

Tsotsos, J. K. (2011). *A computational perspective on visual attention*. MIT Press.

Tsotsos, J. K., Culhane, S. M., Wai, Y. W., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78*(1–2), 507–545.

Tsotsos, J. K., Kotseruba, I., Rasouli, A., & Solbach, M. D. (2018). Visual attention and its intimate links to spatial cognition. *Cognitive Processing, 19*(Suppl 1), 121–130. https://doi.org/10.1007/s10339-018-0881-6

Tsotsos, J. K., & Kruijne, W. (2014). Cognitive programs: Software for attention's executive. *Frontiers in Psychology, 5*, 1260. https://doi.org/10.3389/fpsyg.2014.01260

Ullman, S. (1984). Visual routines. *Cognition, 18*(1–3), 97–159.

Ullman, S. (1996). *High-level vision*. MIT Press.

Ursino, M., & La Cara, G. E. (2004). A model of contextual interactions and contour detection in primary visual cortex. *Neural Networks, 17*(5–6), 719–735. https://doi.org/10.1016/j.neunet.2004.03.007

Usher, M., & Cohen, J. D. (1999). Short term memory and selection processes in a frontal-lobe model. In D. Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience. Perspectives in neural computing.* Springer.

Vatterott, D. B., & Vecera, S. P. (2015). The attentional window configures to object and surface boundaries. *Visual Cognition, 23*(5), 561–576. https://doi.org/gh3q

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling. *Philosphical Transactions of the Royal Society of London. B Biological Sciences, 372*(1714). https://doi.org/gg3qxc

Wagemanas, J. (2014). *Oxford handbook of perceptual organization*. Oxford University Press.

Wang, D. L. (1995). Emergent synchrony in locally coupled neural oscillators. *IEEE Transactions on Neural Networks, 6*(4), 941–948. https://doi.org/10.1109/72.392256

Wang, D. L. (2005). The time dimension for scene analysis. *IEEE Transactions on Neural Networks, 16*(6), 1401–1426. https://doi.org/10.1109/TNN.2005.852235

Wang, D. L., & Terman, D. (1995). Locally excitatory globally inhibitory oscillator network. *IEEE Transactions on Neural Networks, 6*(1), 283–286. https://doi.org/dw82rn

Wang, D. L., & Terman, D. (1997). Image segmentation based on oscillatory correlation. *Neural Computation, 9*(4), 805–836. https://doi.org/10.1162/neco.1997.9.4.805

Wannig, A., Stanisor, L., & Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience, 14*(10), 1243–1244. https://doi.org/10.1038/nn.2910

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences, 7*(12), 553–559. https://doi.org/10.1016/j.tics.2003.10.012

Yuille, A., & Geiger, D. (2003). Winner-take-all networks. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 1228–1231). MIT Press.

Zhang, N. R., & von der Heydt, R. (2010). Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *Journal of Neuroscience, 30*, 6482–6496. https://doi.org/10.1523/JNEUROSCI.5168-09.2010

Zhou, H., Schafer, R. J., & Desimone, R. (2016). Pulvinar-cortex interactions in vision and attention. *Neuron, 89*(1), 209–220. https://doi.org/10.1016/j.neuron.2015.11.034

APPENDIX C

Mogu li Kognicija i Emocije Utjecati na Vid? [Can Cognition and Emotions Affect Vision?]

Marić, M. i Domijan, D. (2018). Mogu li kognicija i emocije utjecati na vid? [Can cognition and emotions affect vision?]. *Psihologijske teme, 27*(2), 311–338. https://doi.org/10.31820/pt.27.2.9

SAŽETAK

U radu su prikazani teorijski argumenti i empirijske potvrde za i protiv ideje da kognitivni procesi (mišljenje, rezoniranje, očekivanje, vjerovanje) ili emocije i motivacija mogu izravno utjecati na i mijenjati sadržaj vida. Prema hipotezi o modularnosti uma i Marrovoj računalnoj teoriji, percepcija je informacijski zatvoren modul s fiksnom, urođenom arhitekturom. Ona se zasniva na specifičnim načelima koja su bitno drugačija od općeg kognitivnog funkcioniranja. Percepcija mora biti kognitivno inpenetrabilna zato jer mora stvoriti točnu mentalnu reprezentaciju vanjskog svijeta i time omogućiti jedinki uspješno snalaženje u njemu. Suprotno tome, prema modelu prediktivnog kodiranja, kao suvremenog oblika zagovaranja penetrabilnosti vida, mozak stalno generira predikcije koje olakšavaju i usmjeravaju procesiranje osjetnih informacija i posljedično mijenjaju ono što vidimo. U novije vrijeme, mnoga bihevioralna i neuroznanstvena istraživanja pokazuju da se vid doista mijenja pod utjecajem naučenih asocijacija i konteksta, kao i socijalne kognicije i emocija, što potvrđuje da je percepcija kognitivno penetrabilna. Međutim, ne slažu se svi s ovim zaključkom budući da su identificirane brojne metodološke i/ili interpretacijske poteškoće koje otežavaju donošenje definitivnog zaključka. Na kraju, opisani su mogući smjerovi za daljnja teorijska i empirijska istraživanja koja bi nas trebala približiti razrješenju ovog složenog pitanja.

Ključne riječi: eksperimentalna metodologija, emocije, kognicija, kognitivna neuroznanost, vid

SUMMARY

The paper reviews theoretical arguments and empirical findings in support and against the idea that cognitive processes (thinking, reasoning, expectations, beliefs) or emotions and motivation can directly influence and change the content of visual perception. According to the modularity of mind hypothesis and Marr's computational theory, perception is informationally encapsulated module with fixed, innate architecture. It is based on a specific set of principles that are much different from general cognitive functioning. Perception must be cognitively impenetrable because its task is to create accurate mental representation of the external environment that will enable individual to successfully navigate through it. In contrast, predictive coding model, as an example of modern advocate of penetrability of vision, assumes that brain constantly generates predictions that facilitates and redirects visual information processing and consequently alters what we see. Recently, many behavioural and brain studies revealed that visual perception is indeed altered under the influence of learned associations and context, as well as social cognition and emotions thus lending support to the claim that perception is cognitively penetrable. However, not all agree with this conclusion because there are many methodological and interpretational pitfalls identified in previous research that prevents reaching consensus. At the end, we described potential avenues for further theoretical and empirical research that will bring us closer to the answer to this perplexing question.

Keywords: cognition, cognitive neuroscience, emotions, experimental methodology, visual perception

1. UVOD

Pretpostavka da je percepcija kognitivno penetrabilna podrazumijeva da različiti kognitivni procesi poput mišljenja, vjerovanja, emocija, potreba i drugih mogu izravno djelovati na vidno procesiranje tako da mijenjaju njegov konačni rezultat. Dakle, ono što mislimo i osjećamo doslovno može utjecati na način na koji vidimo svijet oko sebe (Balcetis, 2016; Vetter i Newen, 2014). U terminima obrade informacija u hijerarhijski organiziranom sustavu kao što je vid, to znači da silazni ili odozgo-prema-dolje (*engl. top-down*) procesi mogu direktno utjecati na uzlazne ili odozdo-prema-gore (*engl. bottom-up*) procese (Palmer, 1999). Međutim, problematična posljedica kognitivne penetrabilnosti jest njezina interferencija s temeljnom zadaćom vida, a to je dostaviti što točniju informaciju o okolini radi lakšeg snalaženja u njoj (Marr, 1982). Iz toga proizlazi da vid nije toliko pouzdan izvor informacija kao što nam sugerira naše svakodnevno iskustvo (Lupyan, 2017b). Stoga je važno teorijsko i empirijsko pitanje postoji li barem dio vida koji je funkcionalno nezavisan od kognicije i emocija. S druge strane, ako su svi aspekti percepcije kognitivno penetrabilni, pitanje je na koji je način naše ponašanje usklađeno s fizikalnim obilježjima vanjskog svijeta. Ova pitanja od velike su važnosti ne samo za psihologiju nego i za filozofiju uma, neuroznanost, psihijatriju, pa čak i estetiku. Koliko je kognitivna penetrabilnost vida aktualno i važno pitanje pokazuje i činjenica da su nedavno dva časopisa *Consciousness and Cognition*[1] i *Frontiers in Psychology*[2] objavili tematske brojeve posvećene ovoj temi. Cilj ovog rada je rasvijetliti prirodu pretpostavke o kognitivnoj penetrabilnosti vida te napraviti pregled teorijskih argumenata za i protiv, kao i empirijskih nalaza koji ih podržavaju. Također, ponudit će se i nove ideje za razrješenje ovog složenog pitanja.

2. NEURONSKE OSNOVE VIDA

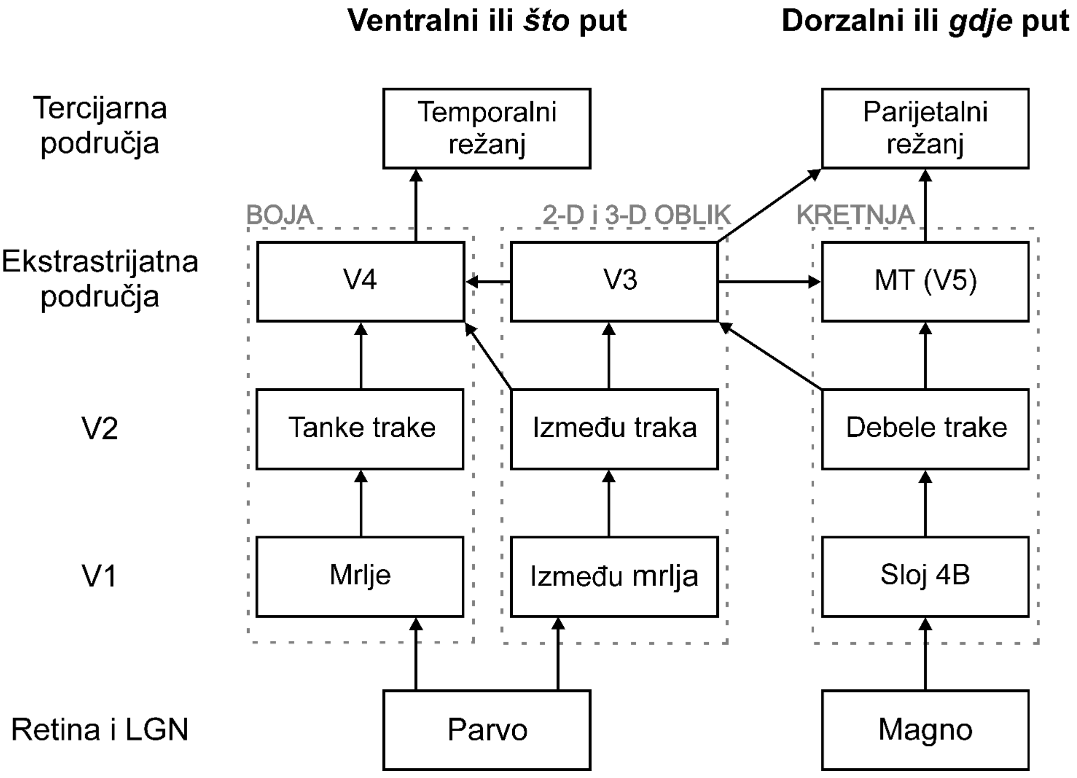
U okviru neuroznanosti, vid je najistraženiji osjetni modalitet. Ova istraživanja pokazala su da vid nije jednostavan proces preslikavanja intenziteta svjetla kojeg registrira retina u unutarnju reprezentaciju u mozgu, već se sastoji od cijelog niza složenih neuronskih mreža

[1] vol. 47, str. 1–112 (2017), *Cognitive penetration and predictive coding*. Dostupno na: http://www.sciencedirect.com/science/journal/10538100/47/supp/C?sdc=1

[2] vol. 8 (2017), *Pre-cueing effects on perception and cognitive penetrability*. Dostupno na: http://journal.frontiersin.org/researchtopic/4600/pre-cueing-effects-on-perception-and-cognitive-penetrability#articles

specijaliziranih za analizu različitih aspekata vidnih podražaja (Bullier, 2001; Lennie, 1998). Zbog toga je u analizi problema kognitivne penetrabilnosti vida potrebno uključiti i neuroznanstvene spoznaje o interakcijama među kortikalnim centrima koji su u njega uključeni.

Obrada vidnih informacija započinje apsorpcijom fotona svjetlosti u fotoreceptorima u retini, a nastavlja se u dva odvojena paralelna puta koji čine parvo i magno stanice. Parvo stanice osjetljive su na podražaje visoke prostorne frekvencije (jer imaju mala receptivna polja), niske temporalne frekvencije, niskog kontrasta i pokazuju oponentne reakcije na boje (crveno-zeleno i plavo-žuto). S druge strane, magno stanice imaju velika receptivna polja, osjetljive su na podražaje niske prostorne frekvencije, visoke temporalne frekvencije, visokog kontrasta i ne pokazuju oponentnost na boje (Callaway, 2005). Parvo i magno stanica iz lateralne koljenaste jezgre (*lateral geniculate nucleus - LGN*) projiciraju u različite slojeve u primarnom vidnom korteksu, koji se još označava i kao V1 ili strijatni korteks ili Brodmanovo područje 17 te predstavlja prvi korak u kortikalnoj obradi vidnih signala. Projekcije iz primarnog vidnog korteksa prema ekstrastrijatnim područjima mogu se podijeliti u dva odvojena funkcionalna sustava ili vidna puta. To su ventralni i dorzalni put koji nastavljaju i dalje elaboriraju supkortikalne parvo i magno puteve. Ventralni put sudjeluje u identifikaciji objekata pa se još naziva i *što* put. S druge strane, dorzalni put ima važnu ulogu u lokalizaciji objekata u prostoru pa se još naziva i *gdje* put (Goodale i Milner, 1992; Ungerleider i Mishkin, 1982). Paralelni vidni putovi prikazani su na Slici 1.



Slika 1. *Shematski prikaz paralelnih vidnih putova. Supkortikalni parvo i magno putovi koji su jasno odvojeni u retini i LGN-u, dalje se u korteksu razdvajaju i elaboriraju. Crni pravokutnici prikazuju anatomske strukture, a sivi iscrtani pravokutnici označavaju funkcionalnu specijalizaciju za boju, oblik i kretnju. Mrlje i trake se odnose na izgled strijatnog korteksa nakon bojenja citokromatskom oksidazom. Prilagođeno prema Coren i sur. (2003).*

Klasičan pristup razumijevanju vida polazi od pretpostavke da opisani vidni putovi čine hijerarhijsku strukturu u kojoj se vid postupno gradi od elementarnih senzornih obilježja koja se nadograđuju i elaboriraju u nizu kortikalnih centara i na kraju dovode do prepoznavanja objekta u anteriornom temporalnom režnju i prostorne lokalizacije objekta u posteriornom parijetalnom režnju (Lennie, 1998). Iz toga proizlazi da je tok informacija u vidnom sustavu jednosmjeran, odnosno implicira striktno uzlazno procesiranje. Međutim, neuroznanstvena istraživanja pokazuju da među vidnim centrima postoje i brojne povratne veze. One prenose široki raspon informacija od usmjeravanja pažnje, očekivanja, vrsta perceptivnog zadatka, radnog pamćenja pa do motornih naloga kojima dinamički mijenjaju odgovore neurona u vidnim putovima (Gilbert i Li, 2013), što upućuje na direktnu interakciju između uzlaznih i silaznih procesa u vidnom sustavu (Bullier, 2001).

Na prvi pogled, čini se da sam pojam povratnih veza podrazumijeva funkcionalnu povezanost vida i kognicije. Međutim, povratne veze imaju drugačija obilježja od uzlaznih veza. Povratne veze najčešće su modulatorne, što znači da mogu mijenjati aktivnost neurona jedino ako je on već aktivan preko praga, a tu aktivaciju su osigurale uzlazne veze (Sherman i Guillery, 2002). Unutar svakog kortikalnog centra, uzlazne i silazne projekcije ostvaruju sinaptičke kontakte u različitom sloju (Callaway, 2004) i ne formiraju snažne povratne petlje (Crick i Koch, 1998). Nadalje, povratne veze ne mogu mijenjati specifično obilježje za koje je neuron osjetljiv već samo pojačavaju ili smanjuju njegovu reakciju na isto obilježje (Martinez-Conde i sur., 1999). Na osnovu opisanih obilježja, Macknik i Martinez-Conde (2009) zaključuju da je glavna uloga povratnih veza održavanje prostorne pažnje na dijelu vidnog polja. Prema Firestoneu i Schollu (2016) za raspravu o interakciji između percepcije i kognicije potrebno je razlikovati mijenjanje procesiranja od mijenjanja ulaza u vid. Na primjer, zatvaranjem očiju ili okretanjem glave možemo kontrolirati ono što doživljavamo vidom, ali tim postupkom mijenjamo samo ulaz u percepciju, ali ne i vidno procesiranje pa se to ne može uzeti kao argument za utjecaj kognicije na percepciju. Slično djeluje i prostorna pažnja budući da selektivno usmjerava obradu ka jednom dijelu vidnog polja te na taj način mijenja ulaz u vid. Zbog ovog se razloga ni njezino djelovanje ne može smatrati kognitivnim utjecajem na vid.

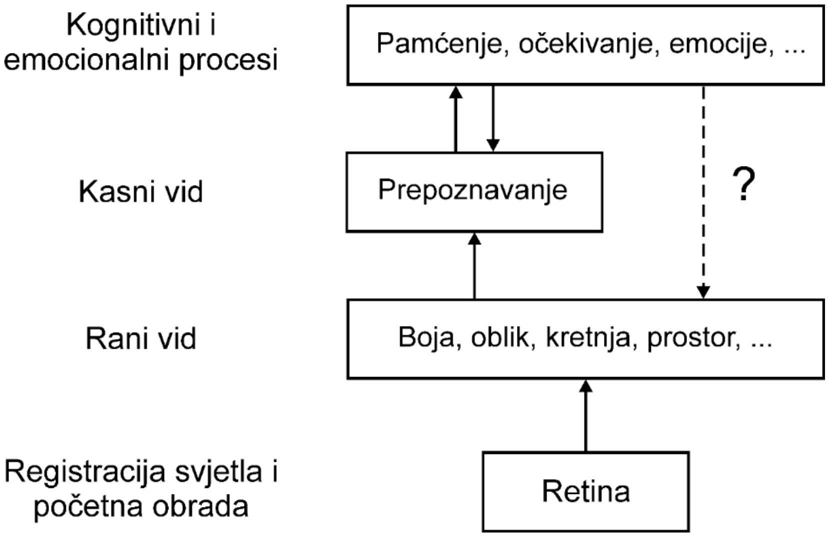
3. VID KAO MODUL

Polazeći od Fodorove (1983) hipoteze o modularnosti uma, Pylyshyn (1999) razvija i brani ideju o kognitivnoj inpenetrabilnosti percepcije. Prema Fodoru (1983), ljudski um sastoji se od središnjeg sustava, senzornih procesora i ulaznih modula. Središnji sustav čine kognitivni procesi kao što su mišljenje, rezoniranje i donošenje odluka. To je kognitivni aparat koji koristimo svjesno i namjerno kako bismo riješili neki problem ili kako bismo razumjeli neku situaciju. Uloga ulaznih modula jest da preuzmu informacije koje registriraju senzorni procesori, odnosno osjetni organi (npr. retina, pužnica, koža i drugi) i preoblikuju ih na način da budu dostupne središnjem sustavu. Temeljna razlika između središnjeg sustava i modula je u tome što su moduli informacijski zatvoreni (enkapsulirani), dok središnji sustav nije. To znači da moduli reagiraju samo na određenu vrstu informacija za koju su specijalizirani, dok središnji sustav može prihvatiti i obraditi bilo kakvu informaciju koja je relevantna za zadatak. Dakle, središnji sustav može primiti informaciju iz svih modula, ali ne može utjecati na module. On nema uvid u to što i kako moduli rade niti može usmjeravati obradu u njima, nego samo dobiva krajnji rezultat njihove obrade. Također, moduli ne mogu stupiti niti u međusobnu interakciju.

Štoviše, obrada informacija u modulima je automatizirana i oni imaju fiksnu neuralnu arhitekturu koja je urođena.

Pylyshyn (1999) smatra da je vidna percepcija modul u Fodorovom smislu te riječi. Iz toga proizlazi da svi kognitivni procesi koji su dio središnjeg sustava ne mogu utjecati na vid te je on stoga kognitivno inpenetrabilan. Preciznije, onaj dio vida koji je kognitivno inpenetrabilan naziva se *ranim vidom.* On obuhvaća procese u rasponu od registracije svjetla na fotoreceptorima, analize različitih perceptivnih obilježja kao što su boja, oblik, pokret, tekstura, dubina pa do stvaranja jedinstvene trodimenzionalne reprezentacije površine objekata u okolini (Adelson i Bergen, 1991; Marr, 1982). Rani vid nije pod utjecajem pažnje pa se još naziva i vid prije pažnje (*engl. pre-attentive vision*) (Julesz i Bergen, 1983; Treisman i Gelade, 1980). Za razliku od ranog vida, *kasni vid* je kognitivno penetrabilan, a sastoji se od prepoznavanja i identifikacije objekata na temelju informacija iz dugoročnog i semantičkog pamćenja, usmjeravanja pažnje i svjesnih pretpostavki o percipiranom subjektu. Dakle, semantičke informacije počinju djelovati na vidno procesiranje tek nakon što je rani vid odradio svoj posao, odnosno odredio reprezentaciju 3-D površine objekata u vidnom polju. Slika 2 shematski prikazuje odnos ranog vida, kasnog vida i kognitivnih procesa. Ovdje treba istaknuti da se pojam vidna percepcija ponekad koristi za skupno označavanje i radnog i kasnog vida a ponekad samo za označavanje ranog vida. Kako bi izbjegli nejasnoće, u daljnjem tekstu ćemo pojam vidna percepcija koristiti samo kao sinonim za rani vid.

Pylyshyn (1999) koncept kognitivne inpenetrabilnosti u velikoj mjeri temelji na Marrovoj (1982) teoriji prema kojoj se vid sastoji od tri odvojene razine obrade koje su posložene u uzlaznu hijerarhiju: prvobitna skica, 2.5-D skica i 3-D model. Pri tome, rani vid čine prvobitna skica i 2.5-D skica. Najprije prvobitna skica pronalazi promjene u intenzitetu svjetla na slici, odnosno detektira rubove. Rezultati obrade u prvobitnoj skici šalju se u 2.5-D skicu. Ona integrira reprezentaciju rubova s još složenijim aspektima podražaja kao što su stereopsis i pokret. Rezultati analize u 2.5-D skici se dalje prenose u 3-D model koji transformira reprezentaciju iz egocentričnog referentnog okvira u alocentrični, odnosno okvir koji ima ishodište na objektu. Drugim riječima, 3-D model osigurava reprezentaciju adekvatnu za prepoznavanje objekata. Iz navedenog opisa vidimo da rezultati obrade u prvobitnoj i 2.5-D skici utječu na prepoznavanje objekata, ali prepoznavanje objekata ne može utjecati na obradu u prvobitnoj skici ili u 2.5-D skici. Zbog toga su ove skice kognitivno inpenetrabilne (Raftopoulos, 2001).


Kognitivni i emocionalni procesi
Pamćenje, očekivanje, emocije, ...
Kasni vid
Prepoznavanje
?
Rani vid
Boja, oblik, kretnja, prostor, ...
Registracija svjetla i početna obrada
Retina

Slika 2. *Shematski prikaz odnosa između ranog vida, kasnog vida te kognitivnih i emocionalnih procesa. Temeljno pitanje preglednog rada je postoji li povratna veza označena iscrtanom strelicom.*

3.1. Argumenti za Kognitivnu Inpenetrabilnost Vida

Prema Pylyshynu (1999), glavni argumenti za kognitivnu inpenetrabilnost percepcije su:

1. Perceptivne iluzije ne nestaju nakon što shvatimo da se radi o iluziji. Znanje da naša percepcija ne odgovara stvarnom stanju u svijetu i dalje ne mijenja sadržaj percepcije. Na primjer, kod Müller-Lyerove iluzije, iako smo svjesni da su dvije linije koje uspoređujemo jednako duge, to ne mijenja našu percepciju da je linija koja završava krakovima prema van duža od linije koja završava krakovima prema unutra.
2. Postoje mnoge pravilnosti u vidu koje ovise samo o vidnom podražaju i koje se automatski ekstrahiraju iz njega. Načela prema kojima funkcionira vid nisu ista kao i načela mišljenja i rezoniranja prema kojima funkcionira središnji sustav. Na primjer, kod percepcije svjetline, Gilchrist i sur. (1999) identificirali su pravilo prema kojem kao bijelu percipiramo onu površinu koja odašilje najveću količinu svjetla u danom podražaju. Ovakvo pravilo specifično je za percepciju svjetlina jer rješava konkretan računalni problem kako preslikati skalu relativnih omjera intenziteta svjetla na apsolutnu skalu svjetlina od bijelog do crnog. Ovo pravilo ne može se primijeniti u drugim domenama i ne predstavlja opće pravilo po kojem može funkcionirati središnji sustav.
3. Neuropsihološka istraživanja pokazuju djelomičnu neovisnost vida od drugih kognitivnih funkcija. Na primjer, ljudi koji imaju vidnu agnoziju, odnosno oštećenje sposobnosti prepoznavanja objekata, nemaju istovremeno i problema s općim

kognitivnim funkcioniranjem (Farah, 1990). S druge strane, ljudi s deficitom u mišljenju i rezoniranju općenito nemaju problema s vidom.

4. Mnogi empirijski nalazi koji se uzimaju kao argument za penetrabilnost vida zapravo se mogu reinterpretirati na način da znanje utječe na neku drugu razinu obrade informacija koja je izvan onoga što se definira kao rani vid. Na primjer, znanje može utjecati na usmjeravanje prostorne pažnje prije početka ranog vida ili može djelovati na procese donošenja odluka kojoj kategoriji pojedini podražaj pripada, što se odvija nakon završetka obrade u ranom vidu. Kako bi se preciznije odredilo na koju razinu obrade informacija djeluje neka eksperimentalna manipulacija, potrebna je mnogo preciznija kontrola relevantnih varijabli, na što su posebno upozorili Firestone i Scholl (2016), a o tome će biti više riječi u odjeljku Metodološki aspekti istraživanja penetrabilnosti vida.

Pylyshynove argumente protiv penetrabilnosti vida dalje je razradio Raftopoulos (2009, 2014). On se posebno fokusirao na neurofiziološka istraživanja o brzini protoka signala kroz vidni sustav kako bi jasno odredio temporalnu granicu između ranog i kasnog vida. Istraživanja mjerenja vremena latencije odgovora neurona u vidnim kortikalnim centrima upućuju na stvaranje vidne reprezentacije podražaja u tri koraka (Lamme, 2003; Lamme i Roelfsema, 2000; Roelfsema, 2005). Prvi korak odvija se unutar prvih 100 ms nakon zadavanja podražaja, kada traje prijenos signala od retine preko vidnih kortikalnih područja sve do inferiornog temporalnog korteksa. Ovaj je prijenos signala nesvjestan i otporan na povratne veze iz viših kortikalnih područja i zato se smatra čistim uzlaznim oblikom prijenosa vidnih informacija. Prvi korak rezultira ekstrahiranjem elementarnih obilježja kao što su boja, orijentacija rubova, pokret, tekstura koji su važni za kasnije prepoznavanje i kategorizaciju. U drugom koraku koji se odvija oko 120 ms nakon zadavanja podražaja počinju djelovati lateralne projekcije neurona unutar istog kortikalnog područja, kao i povratne veze u ranija područja poput primarnog vidnog korteksa. U ovom koraku formiraju se složenije vidne reprezentacije kao što je reprezentacija površina i razdvajanje lika od pozadine. Tek u trećem koraku koji se odvija između 150 i 200 ms nakon zadavanja podražaja signali iz frontalnih i prefrontalnih područja te mnemoničkih krugova u hipokampusu počinju modulirati perceptivno procesiranje u vidnom korteksu. Stoga Raftopoulos (2009, 2014) tvrdi da je tek treći korak vidnog procesiranja kognitivno penetrabilan, dok prva dva koraka čine procesi koji odgovaraju Marrovom konceptu ranog vida i nisu izravno pod utjecajem kognitivnih procesa.

3.1.1. Nesvjesno Zaključivanje Nasuprot Prirodnim Ograničenjima

Temeljni problem percepcije je u tome što nije dovoljno određena samom retinalnom slikom. Naime, ista retinalna slika može biti posljedica različitih situacija u okolini i može dovesti do različitih interpretacija. Dakle, vidni sustav mora kreirati reprezentaciju okoline na osnovu djelomičnih informacija. Ovaj problem je vrlo složen i zahtijeva netrivijalna rješenja. Primjerice, živimo u trodimenzionalnom prostoru u kojem je važno odrediti udaljenost od objekata radi lakšeg snalaženja u prostoru. Međutim, slika na retini je dvodimenzionalna. Stoga vidni sustav mora rekonstruirati treću dimenziju na osnovu spajanja informacija iz dviju retina. Kako vidni sustav rješava ovaj problem? Klasični odgovor ponudio je von Helmholtz (1867/1924) konceptom nesvjesnog zaključivanja prema kojem vid koristi skrivene pretpostavke i znanje o vanjskom svijetu pomoću kojih dolazi do perceptivnih zaključaka o najvjerojatnijem stanju stvari u okolini koje je moglo proizvesti danu retinalnu stimulaciju. Pri tome, vid koristi procese slične zaključivanju i rješavanju problema kako bi sirove retinalne signale pretvorio u percepciju. Ljudi nisu svjesni odvijanja ovih procesa, već samo konačnog rezultata. Ove ideje dalje su razvili i nadogradili Gregory (1970) i Rock (1983) u okviru konstruktivističke škole mišljenja koja predstavlja dominantni teorijski pravac u suvremenoj znanosti o vidu (Palmer, 1999). Jedna od implikacija nesvjesnog zaključivanja je da je vid kognitivno penetrabilan jer je znanje i zaključivanje direktno uključeno u stvaranje vidne reprezentacije. Međutim, Pylyshyn (1999) smatra pogrešnim pripisivati vidu procese zaključivanja i rješavanja problema. Umjesto toga vid treba promatrati kao složeni sustav za obradu informacija koji radi po svojim specifičnim pravilima.

Alternativno objašnjenje kako riješiti problem konstrukcije vidne reprezentacije na osnovu djelomičnih informacija ponudio je Marr (1982). On smatra da u ranom vidu postoje specijalizirani mehanizmi koji utjelovljuju rješenja zasnovana na prirodnim ograničenjima koja postoje u fizikalnom svijetu. Ovi mehanizmi evolucijom su ugrađeni u vidni sustav i aktiviraju se automatski kada gledamo neki podražaj i ne zahtijevaju procese zaključivanja i rješavanja problema. Pri tome, važno je napomenuti da ovi mehanizmi ne jamče uvijek točnu interpretaciju svih podražaja, nego samo onih koji se najčešće javljaju u danim uvjetima u našoj okolini.

Na primjer, Marr i Poggio (1976) predložili su model stereopsisa koji rješava problem korespondentnosti, odnosno pronalazi korespondentne točke na obje retine i izračunava disparitet među njima. To nije jednostavno budući da postoji ogroman broj kombinacija uparivanja, tj. jedna točka u lijevom oku može odgovarati velikom broju točaka u desnom oku.

Kako bi riješili ovaj problem, Marr i Poggio (1976) identificirali su nekoliko prirodnih ograničenja koja odgovaraju karakteristikama okoline, a koja mogu poslužiti vidnom sustavu za lakše pronalaženje parova točaka. Jedno takvo ograničenje je kompatibilnost obilježja prema kojem se mogu uparivati samo slična obilježja jer isti objekt ne može proizvesti bitno drugačiju sliku u oba oka. Nadalje, Marr i Poggio (1976) su pokazali kako se dana ograničenja mogu ugraditi u arhitekturu neuronske mreže u obliku uzoraka ekscitacija i inhibicija neurona. Mreža se sastojala od neurona koji su kodirali različite disparitete točaka u obje retine. Računalne simulacije predložene mreže pokazale su da je uspješno pronalazila korespondentne točke i na taj način odredila dubinu na osnovu dispariteta dvije retinalne slike. Dakle, mreža nije trebala posebno eksterno znanje niti procese zaključivanja i rješavanja problema kako bi uparila točke iz obje retine. Znanje je ugrađeno u samu strukturu mreže poštujući postojeća prirodna ograničenja.

4. KOGNITIVNA PENETRABILNOST VIDA

4.1. Novi Pogled

Koncepti kognitivne inpenetrabilnosti i modularnosti vida u suprotnosti su s eksperimentalnim nalazima nastalima unutar popularnog znanstvenog pokreta poznatog pod nazivom *novi pogled* (Bruner, 1957). Prema ovom gledištu, kognitivni procesi mogu utjecati na percepciju, stoga je ono što vidimo rezultat interakcije naših vjerovanja, želja, emocija, motivacije i jezika sa sirovim osjetnim informacijama. Bruner i Goodman (1947) su u brojnim eksperimentima pokazali da je percepcija pod utjecajem znanja i vjerovanja promatrača o svijetu koji percipira. Primjerice, siromašna djeca sustavno precjenjuju veličinu novčića u odnosu na bogatiju djecu, a gladni ljudi će prije percipirati hranu ili identificirati riječi povezane s hranom od sitih ljudi. Na osnovu ovakvih nalaza zaključili su da vrijednosti i potrebe pojedinca utječu na sve razine hijerarhijske strukture percepcije i tako mijenjaju način na koji percipiramo svijet.

Slično von Helmholtzu, Bruner (1957) smatra da vid određuju dva bitna obilježja: kategorizacija i testiranje hipoteza. Vid se promatra kao proces sličan znanosti u kojem se formuliraju i testiraju hipoteze. Hipoteze se odnose na pripadnost objekta određenoj kategoriji. Ako se hipoteza pokaže pogrešnom, odbacuje se i kreira se nova hipoteza koja se zatim testira s obzirom na dane osjetne podatke. Na taj način, vid se kroz niz ciklusa testiranja i odbacivanja hipoteza postupno približava identitetu objekta, odnosno kategoriji kojoj pripada. Stoga se vid može promatrati i kao oblik rješavanja problema u kojem ulazni signali koji dolaze kroz osjete predstavljaju problem koji treba riješiti. Nedostajući podaci potrebni za rješenje dolaze od silaznih procesa, odnosno znanja, očekivanja, potreba i vjerovanja. Rješenje problema, odnosno krajnji rezultat vida je kategorija kojoj pripada objekt koji se promatra. Iz ovih razmatranja proizlazi da nema razlike između načela po kojima funkcionira vid i načela po kojima funkcionira mišljenje. Popularnost ovog gledišta proizlazi i iz činjenice dobrog slaganja s našom intuicijom, odnosno svakodnevnim iskustvom. Na primjer, kada smo gladni, čak i neki neodređeni, neutralni podražaj možemo percipirati kao hranu. Kada smo uplašeni, i bezazleni podražaj kao što je šuštanje lišća na ulici percipiramo kao prijeteći što je ovjekovječeno u izreci *u strahu su velike oči*. Također, primjer za penetrabilnost su i mađioničarski trikovi u kojima mađioničar manipulira s onim što vidimo tako što kreira lažna očekivanja koja nas iznevjere i zbog toga zabavljaju.

4.2. Teorijski Argumenti za Penetrabilnost Vida

U novije vrijeme pokret novi pogled ponovno je oživio razvojem novih teorijskih konstrukata kao što je prediktivno kodiranje i Bayesijanski modeli percepcije (Clark, 2013; Friston, 2010; Hohwy, 2013; Lupyan, 2015; O'Callaghan, Kveraga, Shine, Adams i Bar, 2017; Vetter i Newen, 2014). Osnovna ideja potekla je od von Helmholtza (1867/1924) koji je predložio model eferentnih kopija koje generira motorni sustav kako bi predvidio senzorne posljedice izvršenja motornih naloga. Na taj način, mozak osigurava stabilnu percepciju položaja objekata nakon izvršenja pokreta očiju (Schubotz, 2015). Svrha prediktivnog kodiranja je skraćivanje vremena i smanjenje količine procesnih resursa potrebnih da se obradi ogromna količina senzornih informacija koje neprestano bombardiraju našu retinu (Huang i Rao, 2011). Koncept prediktivnog kodiranja pretpostavlja da mozak, u svrhu predviđanja kakve će vjerojatno biti nadolazeće senzorne informacije, kontinuirano stvara model vanjskog svijeta i mijenja ga i nadograđuje prema pravilima Bayesijanske statistike (Friston, 2010; Rao i Ballard, 1999).

Model okoline oblikuje se na osnovu prošlih iskustava pohranjenih u pamćenju, kontekstualnih asocijacija, kao i senzornih informacija iz drugih osjetnih modaliteta. Naime, objekte opažamo u tipičnim uvjetima u kojima se često pojavljuju zajedno s drugim objektima, a to nam iskustvo omogućava učenje asocijacija među njima. Asocijacije potom postaju bogat izvor informacija kojima se rukovodimo prilikom identifikacije objekta. U slučaju dvosmislenog podražaja, informacija o kontekstualnom okruženju olakšat će nam njegovu interpretaciju (Bar, 2004).

Anatomska osnova za prediktivno kodiranje su povratni neuronski putovi od anteriornog temporalnog i posteriornog parijetalnog korteksa prema primarnom vidnom korteksu (Angelucci, Levitt, i Lund, 2002; Bullier, 2001). U V1 se uspoređuju silazni signali koji nose informaciju o predikciji s nadolazećim uzlaznim, odnosno senzornim signalima, a njihovo međusobno neslaganje predstavlja pogrešku predikcije. Informacija o pogrešci prenosi se uzlaznim vezama natrag do viših vidnih centara, gdje se prediktivni model mijenja, odnosno ispravlja sve dok se ne smanji pogreška predikcije. Drugim riječima, više razine povratnim vezama pokušavaju predvidjeti odgovore jedinica na nižim razinama u vidnoj hijerarhiji (Rao i Ballard, 1999).

Na staničnoj razini, pretpostavlja se da su piramidalne stanice na površini korteksa nosioci informacije o veličini pogreške predikcije nastale usporedbom silaznih i uzlaznih signala, a da su piramidalne stanice u dubljim slojevima korteksa nosioci silaznih predikcija (Friston, 2008; Mumford, 1992). Nadalje, pretpostavlja se da pažnja omogućuje fleksibilno kontroliranje tolerancije s obzirom na preciznost usporedbi i pogrešaka predikcije. Pri tome, pažnja ostvaruje svoj učinak na neurone putem modulacije neurotransmisije acetilkolinom ili putem brze sinkronizacije oscilatorne aktivnosti (Feldman i Friston, 2010).

Važna implikacija prediktivnog kodiranja jest da je vid interaktivan, dvosmjeran proces nadopunjavanja uzlaznih i silaznih signala. Drugim riječima, mozak nikada ne percipira okolinu kao praznu sliku, nego uvijek kreće od neke pretpostavke o tome kako bi okolina mogla izgledati. Friston (2010) navodi da mozak nastoji što je više moguće smanjiti iznenađenje, odnosno pogrešku predikcije te zbog toga stalno procjenjuje stanje u okolini aktivnim zaključivanjem, odnosno prediktivnim kodovima koji silaznim putovima ponderiraju uzorke ekscitacije u nižim vidnim centrima prema načelima Bayesijanske statistike. Stoga se model prediktivnog kodiranja može smatrati oblikom kognitivne penetrabilnosti vida (Lupyan, 2015; Vetter i Newen, 2014). Međutim, Macpherson (2017) upozorava da je moguće kreirati različite verzije modela prediktivnog kodiranja, među kojima i one koji ne podrazumijevaju penetrabilnost vida.

4.3. Neuroznanstveni Argumenti za Penetrabilnost Vida

Istraživanja funkcionalnim oslikavanjem mozga putem magnetske rezonance (fMRI) pokazala su da znanje o objektu može kreirati očekivanja koja mijenjaju aktivnost u vidnim centrima koji reprezentiraju elementarna perceptivna obilježja kao što su boja i oblik. Vienbroucke, Fahrenfort, Meuwese, Scholte i Lamme (2016) prikazivali su ispitanicima nijansu žute boje koja se po svom tonalitetu nalazi na pola puta između crvene i zelene. Žuta boja prikazana je ili na objektu koji je tipično crven (npr. rajčica), objektu koji je tipično zelen (npr. krastavac) ili potpuno nepoznatom objektu. Koristeći tehniku dekodiranja neuronskih signala, pokazali su da mogu predvidjeti gledaju li ispitanici žutu boju na tipično crvenom ili tipično zelenom objektu samo na osnovu neuronskog odgovora na čistu crvenu ili zelenu boju u kortikalnom području V4 za koje se zna da ima važnu ulogu u reprezentaciji boja. Drugim riječima, neuronski odgovor na žutu boju bio je pomaknut prema crvenoj boji kada su je ispitanici gledali na rajčici ili prema zelenoj boji kada su je gledali na krastavcu. Slično tome, Kok, Brouwer, van Gerven i de Lange (2013) pokazali su da kreiranjem očekivanja o smjeru kretanja skupine točkica dolazi do promjene aktivnosti u ranim razinama vidne hijerarhije koja uključuju V1 i V2. Zanimljivo je da nije došlo do modulacije aktivnosti u V5 (MT) području, što bi se očekivalo s obzirom na spoznaje o reprezentaciji kretanja upravo u tom području.

Socijalno i emocionalno znanje također može penetrirati u vid na što upućuju EEG i fMRI istraživanja percepcije lica, koja ima ključnu ulogu u socijalnim interakcijama. Kada ispitanici gledaju lica pripadnika iste socijalne grupe kojoj i sami pripadaju (iste rase ili arbitrarno formirane grupe) dolazi do pojačane aktivacije u fuziformnom girusu nego kada gledaju lica pripadnika druge grupe (Golby, Gabrieli, Chiao i Eberhardt, 2001; Van Bavel, Packer i Cunningham, 2008). U istom smjeru, percepcija lica iz iste grupe dovodi do pojačanog N170 vala koji se smatra najranijom komponentom vidnog procesiranja lica (Ratner i Amodio, 2013). Nadalje, afektivna vrijednost podražaja utječe na vid vrlo rano, što se vidi iz moduliranja ERP signala u rasponu od 120 do 160 ms od zadavanja podražaja (Olofsson, Nordin, Sequeira i Polich, 2008). Ovo ima veliku važnost za preživljavanje jer signalizira jedinki prilaženje objektu ili njegovo izbjegavanje (Barrett i Bliss-Moreau, 2009). Kveraga i sur. (2015) su pokazali da emocionalne predikcije imaju važnu ulogu u brzom prepoznavanju prijetnje. Naime, otkrili su da se bihevioralni odgovor, kao i neuronska reakcija na prijetnju mijenja ovisno o tome radi li se o direktnoj ili indirektnoj prijetnji (gleda li prijeteća životinja direktno u opažača ili ne), što ukazuje na zaključak o integraciji afektivne informacije s vidnom u cilju što brže i točnije interpretacije prijeteće situacije.

Kao glavni izvor silaznih signala, odnosno predikcija identificiran je orbitofrontalni korteks. To je dio frontalnog režnja koji je povezan s nizom važnih funkcija, od ponašajne inhibicije i emocionalne regulacije pa sve do reprezentacije potkrepljenja (nagrade) i donošenja odluka (O'Callaghan i sur., 2017). Njegova uloga u prepoznavanju objekata istraživana je selektivnom stimulacijom ventralnog ili dorzalnog puta. Podražaji niske prostorne frekvencije primarno aktiviraju dorzalni put koji brzo aktivira orbitofrontalni korteks, a koji zatim šalje povratne silazne signale prema fuziformnom girusu u ventralnom putu. S druge strane, podražaji visoke prostorne frekvencije direktno stimuliraju fuziformni girus uzlaznim vezama unutar ventralnog puta (Kveraga, Boshyan i Bar, 2007). Sličnu brzu aktivaciju orbitofrontalnog korteksa koja prethodi aktivaciji temporalnog režnja, odnosno ventralnog puta pokazala su istraživanja u kojima su ispitanici procjenjivali koherentnost degradiranog podražaja (Horr, Braun i Volz, 2014) ili razlikovali lica od ne-lica (Summerfield, Lepsien, Gitelman, Mesulam i Nobre, 2006). Ovi rezultati upućuju na zaključak da orbitofrontalni korteks daje grubu inicijalnu procjenu kategorije kojoj objekt pripada, a ta procjena, putem silaznih veza, usmjerava i ubrzava prepoznavanje objekta u temporalnom režnju. Također, orbitofrontalni korteks igra važnu ulogu i u brzoj obradi afektivne vrijednosti podražaja (Pessoa i Adolphs, 2010).

O'Callaghan i sur. (2017) smatraju da su i vidne halucinacije primjer za kognitivnu penetrabilnosti ranog vida. Analiza sadržaja halucinacija pokazuje da se u njima često javljaju poznati ljudi, lica, objekti ili kućni ljubimci (Barnes i David, 2001). Nadalje, učestalost i jačina halucinacija povezana je s raspoloženjem i fiziološkim stanjem kao što je stres ili umor (Waters i sur., 2014). U okviru modela prediktivnog kodiranja, halucinacije su rezultat pretjerane aktivacije kontekstualnih asocijacija, autobiografskih sjećanja i emocionalnih stanja koja pomiču balans između uzlaznih i silaznih procesa na način da daju veći ponder silaznim procesima. Drugim riječima, internalno generirane predikcije kod halucinacija dominiraju vidom usprkos suprotnim senzornim signalima. Potvrdu za ovu pretpostavku daju istraživanja koja pokazuju abnormalno pojačanu aktivnost upravo u kortikalnim centrima poput orbitofrontalnog korteksa koji predstavljaju izvor silaznih signala tijekom normalne vidne percepcije (Shine, O'Callaghan, Halliday i Lewis, 2014).

4.4. Bihevioralni Nalazi za Penetrabilnost Vida

U zadnjih dvadesetak godina objavljen je veliki broj istraživanja koja ukazuju na kognitivnu penetrabilnost vida. Detaljan popis ovih studija dali su Firestone i Scholl (2016). U nastavku smo odabrali i detaljnije prikazali nekoliko studija koje su izazvale najviše polemika te koje su motivirale raspravu o metodološkim i interpretacijskim problemima, o kojima će biti više riječi u zasebnom odjeljku.

Bhalla i Proffitt (1999) ispitivali su odnos percepcije nagiba brda i fiziološkog potencijala pojedinca. U četiri eksperimenta pokazali su da se brda čine strmijima kada su ljudi: opterećeni nošenjem teškog ruksaka, umorni nakon dugog trčanja, slabe tjelesne kondicije i stariji ili lošijeg zdravlja. U eksperimentu s nošenjem teškog ruksaka ispitanici su davali tri vrste procjene nagiba brda: verbalne, vidne i taktilne. Zanimljiv nalaz je da se verbalno i vidno precjenjivanje nagiba brda povećavalo s opadanjem fiziološkog potencijala zbog opterećenja ruksakom, ali su taktilne procjene ostale točne. Bhalla i Proffitt (1999) su iz ovoga zaključili da je motorički odgovor nesvjestan i daje točnije procjene nagiba, čak i nakon verbalnih uputa o kutu dok ga ispitanik gleda.

Witt, Proffitt i Epstein (2004) su proveli nekoliko eksperimenata u kojima su ispitivali ovisi li efekt fizičkog napora na percipiranu udaljenost predmeta o predviđanju ispitanikove uspješnosti o vlastitoj izvedbi zadatka. Dobiveni rezultati sugeriraju da je percipirana udaljenost u funkciji stvarne udaljenosti koja je određena optičkim varijablama, ali i namjerom za akcijom i naporom povezanim s tom namjerom. Tako, primjerice, bacanje teške lopte na metu u odnosu na laku loptu povećava procjene udaljenosti te mete. Nadalje, pokazalo se da konzumiranje glukoze nakon gladovanja dovodi do toga da se nagib brda percipira manjim (Schnall, Zadra i Proffitt, 2010), a da se udaljenosti percipiraju kraćima (Cole, Balcetis i Dunning, 2013) nego nakon konzumacije nekaloričnog sladila. Isto tako, negativno raspoloženje čini brdo strmijim u odnosu na pozivno raspoloženje (Riener, Stefanucci, Proffitt i Clore, 2011).

Kognitivni i emocionalni utjecaji na percepciju nisu ograničeni samo na percepciju spacijalnih obilježja kao što su nagib ili udaljenost, nego su otkriveni i u području percepcije svjetline i boje. Goldstone (1995) je ispitivao utjecaj kategorizacije na percepciju boje. Ispitanici su najprije učili povezati brojke i slova s određenom bojom u rasponu od crvene do ljubičaste. Pri tome, slova su sustavno prikazivane u crvenijim tonalitetima, a brojke u plavljim (ili obrnuto za drugu grupu ispitanika). Ključni dio eksperimenta je bio u tome da je jedno slovo (E) i jedna brojka (osam) uvijek prikazivana u istoj boji koja je po svom tonalitetu bila negdje na pola puta između crvene i ljubičaste. U drugom dijelu istraživanja ispitanici su podešavali boju brojki i slova. Pokazalo se da je kategorija objekta utjecala na prosudbu boje. Ispitanici su slovo E sustavno procjenjivali crvenijim nego brojku osam (plavlje za drugu grupu ispitanika).

Banerjee, Chatterjee i Sinha (2012) ispitivali su može li dosjećanje (ne)etičkih djela utjecati na percepciju svjetline. Zadatak ispitanika bio je prisjetiti se (ne)etičnih djela iz vlastite prošlosti i detaljno ih opisati uz prateće emocionalne doživljaje. Nakon toga trebali su na numeričkoj skali od jedan do sedam procijeniti stupanj svjetline sobe u kojoj su boravili. Rezultati su pokazali da su ispitanici koji su opisali neetična djela sobu procijenili tamnijom u odnosu na ispitanike koje su opisali etička djela. Slično tome, Song, Vonasch, Meier i Bargh (2012) su ispitivali procjenjuju li ljudi nasmiješena lica svjetlijima u odnosu na namrštena lica. Pretpostavili su da metaforički izrazi poput *svijetli osmjeh*, koji postoji u mnogim jezicima kao što su engleski, njemački, talijanski, kineski ili ruski, imaju podlogu u konceptualnom preslikavanju facijalnih ekspresija na percepciju svjetline. Drugim riječima, pretpostavili su da je osmjeh, odnosno pozitivno raspoloženje povezano sa svijetlim nijansama, dok je mrštenje, odnosno negativno raspoloženje povezano s tamnijim nijansama boje. Ispitanicima su prezentirana po dva shematska lica (smajlića), jedno nasmiješeno i jedno namršteno, pri čemu su oba lica bila izjednačena po svjetlini. U jednom eksperimentu zadatak ispitanika je bio da procijene koje je lice svjetlije, a u drugom da daju apsolutnu procjenu svjetline na numeričkoj skali. Rezultati su pokazali da ispitanici nasmiješeno lice procjenjuju svjetlijim od namrštenog lica neovisno o vrsti procjene koju daju. Isti efekt dobiven je i sa stvarnim licima čime je povećana ekološka valjanost ovog nalaza. Štoviše, razlika u procjenama svjetline je bila veća za stvarna lica nego za shematska lica.

5. METODOLOŠKI ASPEKTI ISTRAŽIVANJA PENETRABILNOSTI VIDA

Problem s originalnim istraživanjima Brunera i njegovih suradnika, koja su inspirirala pokret novi pogled u percepciji, je u tome što nisu kasnije replicirana. Na primjer, pokazalo se da utjecaj socioekonomskog statusa na procjenu veličine novčića postoji samo kada se procjena daje na osnovu dosjećanja, a ne kada se on direktno gleda. Nadalje, neki vrijedni objekti ili simboli nisu izazivali isti efekt kao novčići u originalnoj studiji. Također, otkriven je velik broj moderirajućih varijabli, što je sve ukazivalo na nedovoljnu kontrolu i nedostatnu metodološku rigoroznost (Firestone i Scholl, 2016).

Firestone i Scholl (2016) smatraju da se slični problemi javljaju i u suvremenim istraživanjima. Napravili su detaljnu analizu metodologije i interpretacije rezultata istraživanja koja idu u prilog penetrabilnosti vida te su zaključili da se isti efekti mogu objasniti i bez pozivanja na utjecaj kognicije ili emocija. Najvažniji doprinos njihovog rada je u isticanju zahtjeva za poboljšanjem znanstvene metodologije i kontrole u provedbi istraživanja. Konkretno, izdvojili su šest metodoloških i interpretacijskih poteškoća koja otežavaju donošenje suda o kognitivnoj penetrabilnosti vida i koje će morati uzeti u obzir sva buduća istraživanja. To su: istraživačka strategija usmjerena na potvrđivanje hipoteza, percepcija nasuprot prosudbe, zahtjevi eksperimenta i pristranost pri odgovaranju, razlike u elementarnim obilježjima, periferni efekti pažnje i pamćenje nasuprot prepoznavanju.

5.1. Istraživačka Strategija Usmjerena na Potvrđivanje Hipoteza

Eksperimentalna hipoteza može se provjeriti na dva načina: pronalaženjem efekta koji je u skladu s teorijom i odsustvom efekta kada teorija predviđa njegovu odsutnost. U većini istraživanja o kognitivnoj penetrabilnosti korišten je samo prvi pristup. Stoga, Firestone i Scholl (2016) smatraju da nedostaje mnogo potencijalno važnih nalaza koji mogu pomoći u razrješenju debate o kognitivnoj (in)penetrabilnosti vida. Tek u novije vrijeme počeo se koristiti i drugi način provjere suprotstavljenih teorija testiranjem *jedinstvenih opovrgavajućih hipoteza*.

Konkretan primjer razlikovanja potvrđujućih i opovrgavajućih hipoteza vidljiv je u istraživanjima Firestonea i Scholla (2014) koji su bili inspirirani zanimljivom zgodom iz povijesti umjetnosti poznatom pod nazivom *El Grecova pogreška*. El Greco bio je slavni španjolski slikar iz razdoblja manirizma koji je slikao neobično izduljene ljudske figure. Analizirajući njegove slike te kako bi pokušali objasnili izobličenja likova, povjesničari umjetnosti krivo su pretpostavili da je El Greco bolovao od posebno jakog oblika astigmatizma. Radi se o urođenoj deformaciji rožnice zbog koje je cijeli svijet percipirao izduljeno pa je stoga tako i slikao. Međutim, nekoliko istraživača je argumentiralo da se radi o pogrešnom zaključivanju i da El Greco najvjerojatnije nije svijet doživljavao na izduljen način. Naime, u tom bi slučaju i njegova slikarska platna trebala biti izduljena pa zapravo tragovi pretpostavljenih distorzija ne bi trebali biti vidljivi na samim slikama (Anstis, 2002; Firestone, 2013). Slična pogreška javlja se i prilikom interpretacije nalaza o silaznim procesima. Stoga Firestone i Scholl (2016) preporučaju da se u budućim istraživanjima podjednako koriste i potvrđujuće i opovrgavajuće hipoteze.

U replikaciji istraživanja o utjecaju negativnog ili pozitivnog raspoloženja na percepciju svjetline, Firestone i Scholl (2014) su tražili od ispitanika da procjenu svjetline sobe daju na skali nijansi sivih tonova, a ne na numeričkoj skali. Pri tome, očekivali su da će se utjecaj raspoloženja na percepciju svjetline poništiti. Ako svijet zaista izgleda tamnije (ili svjetlije) nakon indukcije negativnog (ili pozitivnog) raspoloženja, tada će i skala s nijansama sive boje također potamniti (ili posvijetliti) do istog nivoa kao i soba pa će ispitanici birati istu nijansu sive, odnosno neće biti razlike u procjenama svjetline u uvjetu pozitivnog i negativnog raspoloženja. Međutim, razlika u procjenama svjetline pojavila se i u ovom slučaju. Dakle, efekt se pojavio čak i u situaciji kada nije trebao, što ukazuje na važnost formuliranja i ispitivanja jedinstvenih opovrgavajućih predviđanja. Drugim riječima, efekt koji su dobili Banerjee i sur. (2012) ne može biti posljedica utjecaja emocija na percepciju, nego na neki drugi proces. U istom smjeru, Firestone i Scholl (2014) smatraju da su i rezultati istraživanja utjecaja kategorizacije na percepciju boje koje je dobio Goldstone (1995) također posljedica El Grecove pogreške. Naime, uočili su da je u njegovom istraživanju podražaj čiju su boju ispitanici podešavali zapravo bio kopija ciljnog slova. Dakle, ispitanici su istovremeno gledali dva podražaja iste boje. Ako je kategorijalno znanje djelovalo na jedan podražaj, trebalo je imati isti utjecaj i na drugi pa bi se efekti opet trebali međusobno isključiti. Budući da je razlika u procjeni boje ipak dobivena, ona ukazuje na to da nije riječ o perceptivnom efektu kao i u prethodnom istraživanju.

5.2. Percepcija Nasuprot Prosudbe

Budući da je teško odrediti granicu između kognicije i vida, često nije jasno je li ono što vidimo posljedica nekog kognitivnog stanja ili su naši zaključci ili mišljenja posljedica onoga što vidimo. Npr. možemo percipirati boju ili veličinu nekog objekta (cipele), ali možemo samo prosuđivati o njegovoj skupoći, udobnosti i sl. Iz ovog je razloga moguće da su rezultati u istraživanjima o kognitivnoj penetrabilnosti posljedica djelovanja silaznih procesa na prosudbu, a ne na vid. Stoga će buduća istraživanja morati uzeti u obzir razliku između vida i prosudbe te ih empirijski razdvojiti kako bi se izbjegle krive interpretacije nalaza.

Kao primjer ovog problema mogu se navesti rezultati koje su dobili Witt i sur. (2004) o utjecaju fizičkog napora na percepciju udaljenosti mete. Alternativno objašnjenje ovog nalaza jest da su ispitanici prilikom davanja odgovora jednostavno uzeli u obzir i neperceptivne faktore kao što je težina lopte i da nije došlo do stvarne promjene u percepciji udaljenosti. U skladu s tom interpretacijom, Woods, Philbeck i Danoff (2009) su odvojili vidnu percepciju od prosudbe tako da su ispitanicima dali detaljne upute o tome kako da daju svoje odgovore. Naime, ispitanici su morali odgovoriti koliko daleko im objekt stvarno izgleda i da se pri tome potrude isključiti sve druge stvari zbog kojih imaju osjećaj da se meta nalazi na nekoj drugoj udaljenosti te da se potrude zanemariti taj osjećaj udaljenosti mete. Pokazalo se da se uz ovakvu uputu efekt težine lopte gubi, što znači da su rezultati koje su dobili Witt i sur. (2004) ipak odražavali mišljenje odnosno procjenu ispitanika o udaljenosti, a ne kako su tu udaljenost doista percipirali. Međutim, Schnall (2017) smatra da Firestone i Scholl (2016) ne uzimaju u obzir ključne procese koji oblikuju prosudbu: nemogućnost čovjeka da introspektivno pristupi razlozima u podlozi svojih prosudbi, dinamiku razgovora u kontekstu eksperimentalnih istraživanja i pogrešne atribucije. Prema njoj perceptivni procesi funkcioniraju kao i procesi prosudbe budući da nisu dostupni svijesti i služe adaptivnom funkcioniranju.

5.3. Zahtjevi Eksperimenta i Pristranost pri Odgovaranju

Zahtjevi zadatka predstavljaju važnu prijetnju unutrašnjoj valjanosti eksperimentalnih istraživanja (Shaughnessy, Zechmeister i Zechmeister, 2012). Naime, sva istraživanja u psihologiji provode se u socijalnom okruženju te uvijek postoji određena interakcija između eksperimentatora i ispitanika. Pri tome, ispitanici nisu pasivni sudionici, nego se aktivno trude shvatiti pravu svrhu eksperimentalne manipulacije. Također, trude se biti *dobri* ispitanici, odnosno usklađuju svoje ponašanje i odgovore s onim što oni misle da je prava svrha istraživanja kako bi ugodili eksperimentatoru. Zbog toga je potrebno uložiti veliki trud kako bi se sakrili svi znakovi koji ukazuju na hipotezu istraživanja. Klein i sur. (2012) upozoravaju da se u novijim istraživanjima sve manje pažnje posvećuje ovom pitanju, što otežava otkivanje stvarnih efekata i doprinosi problemu replikabilnosti psihologijskih istraživanja.

Kao primjer ovog problema možemo navesti istraživanje koje su proveli Bhalla i Proffitt (1999) o utjecaju fizičkog napora na percepciju nagiba brda. Naime, moguće je da ispitanici ne doživljavaju brdo strmijim zbog nošenja teškog ruksaka, nego prilagođavaju svoje odgovore kako bi udovoljili očekivanjima eksperimentatora. U skladu s tom interpretacijom, Durgin i sur. (2009) su pokazali da je efekt ruksaka nestao u situaciji kada je ispitanicima dana uvjerljiva (ali lažna) prikrivena priča o svrsi nošenja ruksaka. Konkretnije, jednoj grupi ispitanika je objašnjeno da u ruksaku nose elektromiografsku opremu koja služi za nadgledanje mišića gležnja. U tom uvjetu, procjene nagiba brda nisu se razlikovale od kontrolne grupe koja nije nosila ruksak. Nadalje, ispitanici su nakon mjerenja dali svoje mišljenje o provedenom istraživanju te je dio ispitanika naveo očekivanje da će nošenje ruksaka promijeniti percepciju nagiba. Dodatna je analiza pokazala da se efekt ruksaka pojavio upravo kod onih ispitanika koji su bili svjesni hipoteze istraživanja. U sljedećem istraživanju, Durgin, Klein, Spiegel, Strawser i Williams (2012) najprije su tražili od ispitanika da daju procjene nagiba brda, a zatim su im postavili nekoliko pitanja o tome koliko je težak ruksak te da li je ruksak utjecao na njihovu procjenu nagiba. Nakon toga, ponovno su tražili od ispitanika da daju procjene nagiba brda i pokazalo se da se efekt ruksaka izgubio nakon druge procjene. Iz toga možemo zaključiti da su dobro osmišljene prikrivene priče nužne u istraživanjima kognitivne penetrabilnosti vida te da je obvezno provesti debrifing nakon provedenog istraživanja, odnosno pitati ispitanike što misle o eksperimentu kako bi se utvrdilo jesu li bili svjesni njegove prave svrhe.

5.4. Razlike u Elementarnim Obilježjima

Pri dizajnu istraživanja potrebno je obratiti pažnju i na manipulaciju podražajima u eksperimentalnim uvjetima. Npr. utjecaj uzbuđenja na spacijalnu percepciju možemo provjeriti usporedbom grupa visoke i niske razine uzbuđenja na percepciju istog podražaja ili možemo izmjeriti kako ispitanici percipiraju udaljenost uzbudljivog i neuzbudljivog podražaja. Iako oba pristupa imaju prednosti i nedostatke, jedna od poteškoća pri manipulaciji podražajima u eksperimentalnim uvjetima je mogućnost da je namjeravana silazna manipulacija pomiješana s promjenama u fizičkoj (uzlaznoj) razini obilježja podražaja i da te suptilne razlike na uzlaznoj razini mogu biti odgovorne za izmjerene razlike u eksperimentalnim uvjetima. Firestone i Scholl (2016) smatraju važnim razdvojiti djelovanje silaznih od uzlaznih varijabli na način da se primjeni silazni faktor, a eliminira uzlazni faktor, ili obrnuto. Primjerice, Levin i Banaji (2006) su pokazali da lica Afroamerikanaca izgledaju tamnije od bijelih lica, čak i kada se izjednače po prosječnoj svjetlini što ukazuje na djelovanje silaznog faktora znanja o rasi. S druge strane, Firestone i Scholl (2016) su replicirali ovo istraživanje sa zamagljenim verzijama lica kako bi se eliminirala informacija o rasi, a istaknule fizičke razlike na uzlaznoj razini. Pokazalo se da je nakon zamagljivanja većina ispitanika izjavila da su oba lica iste rase (pa čak i da je riječ o istoj osobi), ali su i dalje zamagljeno crno lice procjenjivali tamnijim od zamagljenog bijelog lica, što ukazuje na djelovanje neke uzlazne varijable, a ne silazne varijable kao što je rasa. Međutim, Baker i Levin (2016) su zaključili da su Firestone i Scholl podcijenili sposobnost ispitanika da detektira rasu na zamagljenoj slici budući da je pitanje o rasi bilo otvorenog tipa. Kada se umjesto pitanja otvorenog tipa koristi metoda prisilnog izbora pokazalo se da je 75 – 80% ispitanika točno odredilo rasu. Stoga su potrebna daljnja istraživanja kako bi se razriješilo pitanje mogu li i u kojim uvjetima ispitanici detektirati rasu na zamagljenim slikama. U svakom slučaju, sva buduća istraživanja morat će posvetiti mnogo više pažnje kontroli fizičkih karakteristika podražaja.

5.5. Efekti Usmjeravanja Pažnje

Selektivna pažnja je blisko povezana s vidom jer usmjeravanjem pažnje možemo promijeniti ono što vidimo, pa čak i naglasiti neke objekte i učiniti ih jasnijima. Mijenjanje onoga što vidimo na način da selektivno odabiremo različite lokacije slično je mijenjanju onoga što vidimo pokretima očiju jer u oba slučaja mi odabiremo ulaz u vid. Međutim, u oba slučaja utjecaj pažnje je nezavisan od razloga za usmjeravanje pažnje, odnosno pažnja nije osjetljiva na sadržaj te namjere ili vjerovanja pa se takav utjecaj ne može smatrati kognitivnom penetracijom vida (Firestone i Scholl, 2016). S druge strane, Lupyan (2017a) smatra da neki drugi oblici pažnje, kao što su pažnja usmjerena na obilježja ili na semantičke kategorije doista mijenjaju sadržaj vida na način da mijenjaju izgled objekta. Pri tome je ponudio nekoliko demonstracija u kojima dodatna informacija o tome što se nalazi na slikama bitno mijenja percepciju slike (iako bi zagovornici teze o inpenetrabilnosti vida rekli da mijenja interpretaciju, a ne percepciju slike).

Stoga je važno u budućim istraživanjima kontrolirati ili izmjeriti efekt usmjeravanja spacijalne pažnje kako bi ga se razlikovalo od drugih silaznih procesa. Također, važno će biti razlučiti ima li pažnja usmjerena na obilježja ili na semantičke kategorije drugačiji utjecaj na vid u odnosu na spacijalnu pažnju. Na primjer, istraživanja o dvosmislenim slikama (npr. Neckerova kocka ili slika patka-zec) pokazala su da ispitanici mogu voljnom kontrolom odlučiti koju će od dvije interpretacije percipirati, što bi se moglo interpretirati kao argument za penetrabilnost vida (Churchland, 1988). Naknadna istraživanja pokazala su da se ovaj efekt zapravo svodi na usmjeravanje spacijalne pažnje na različite dijelove slike koji dobivaju prednost u obradi i rezultiraju preferiranjem jedne interpretacije (Long i Toppino, 2004). S druge strane, ispitanici izvještavaju o dužem zadržavanju one interpretacije koja im je poznatija. Razlika u dužini zadržavanja jedne interpretacije nestaje ako se podražaj zarotira za 180° (Peterson, 1994; Peterson i Gibson, 1994; Peterson, Harvey i Weidenbacher, 1991; Vecera i Farah, 1997). Ova istraživanja upućuju na zaključak da prepoznavanje objekta, odnosno prethodno znanje može direktno usmjeravati proces razdvajanje lika od pozadine koji se smatra dijelom ranog vida. Buduća istraživanja moraju razjasniti je li i ovaj efekt zapravo posredovan usmjeravanjem spacijalne pažnje ili predstavlja dokaz za kognitivnu penetrabilnost vida.

5.6. Pamćenje i Prepoznavanje

U mnogim se istraživanjima efekt kognitivne penetracije miješa s prepoznavanjem podražaja. Npr. pridjeljivanje jezičnih oznaka jednostavnim oblicima ubrzava vrijeme vidne pretrage i drugih zadataka prepoznavanja. Međutim, prepoznavanje osim vidnog procesiranja uključuje i pamćenje kako bi se dani vidni podražaj usporedio sa zapamćenom reprezentacijom. Budući da vidno prepoznavanje uključuje i percepciju i pamćenje, preporuka za buduća istraživanja je razlikovati ova dva procesa (jer efekti pamćenja nemaju implikacije za prirodu percepcije) ukoliko želimo govoriti o efektima kognitivne penetracije u vid (Firestone i Scholl, 2016).

Gantman i Van Bavel (2014) su ispitivali efekt vidnog iskakanja moralno važnih riječi (*engl. moral pop-out*). Ispitanicima su tahistoskopski prezentirane smislene i besmislene riječi za koje su morali odlučiti u koju od te dvije kategorije pripadaju (zadatak leksičke odluke). Neke su riječi bile moralno važne (npr. ilegalan), a neke moralno nevažne (npr. ograničen). Rezultati su pokazali da ispitanici točnije identificiraju moralno važne od moralno nevažnih riječi. Međutim, moralno važne riječi su bile međusobno semantički povezane. Zbog toga je moguće da su moralne riječi stvorile pripremu jedna za drugu i da su se lakše prepoznavale zbog olakšanog doziva iz semantičkog pamćenja, a ne zbog utjecaja na vid. Stoga se utjecaj semantičke pripreme (*engl. priming*) ne može smatrati primjerom kognitivne penetrabilnosti vida budući da bilo koja kategorija riječi može u sličnim uvjetima biti lakše detektirana. Upravo to su demonstrirali Firestone i Scholl (2015) u eksperimentu u kojem su dobili isti efekt, odnosno točnije detekcije riječi koristeći kao podražaje riječi povezane s trivijalnom arbitrarnom kategorijom kao što je odijevanje. U drugom eksperimentu koristili su riječi povezane s prometom i opet su dobili isti efekt. Dobiveni rezultati sugeriraju da je semantička povezanost riječi ključan faktor u podlozi efekta kojeg su dobili Gantman i Van Bavel (2014). Stoga se može zaključiti da pamćenje, a ne vidna percepcija, poboljšava detekciju riječi povezanih s moralom (kao i s bilo kojom drugom kategorijom riječi).

6. ZAKLJUČAK

Na osnovu prikazanih teorijskih argumenata kao i empirijskih podataka koji im idu prilog može se jedino zaključiti da nije postignut konsenzus o tome postoji li utjecaj kognicije i emocija na vid. U teorijskom smislu, važno je razjasniti ulogu silaznog procesiranja i povratnih veza u vidnom korteksu. Prediktivno kodiranje i Bayesijanski modeli nisu jedini računalni modeli vidne percepcije. Alternativni pristup razumijevanju uloge silaznih procesa pružio je Grossberg (2013) u okviru teorije adaptivne rezonance. Grossberg (1999, 2017) je predložio hipotezu prema kojoj su svjesna perceptivna stanja rezultat rezonance, odnosno slaganja između uzlaznih i silaznih signala. Kada slaganja nema, silazni utjecaji se brišu, odnosno inhibiraju te je daljnje procesiranje vođeno isključivo uzlaznim signalima. Iz toga proizlazi da je vid zaštićen od kognitivnog utjecaja. Međutim, nije jasno kako bi ova teorija objasnila nalaze koji se navode u prilog kognitivnoj penetrabilnosti vida. Stoga su potrebna daljnja istraživanja, posebno simulacijske studije, koje bi objasnile na kojem nivou obrade informacija u okviru teorije adaptivne rezonance dolazi do djelovanja kognicije i emocija.

U budućim empirijskim istraživanjima potrebno je uvažiti preporuke koje su predložili Firestone i Scholl (2016) kako bi se razlučilo stvarno djelovanje kognicije i emocija na vid od djelovanja drugih (ometajućih) faktora koji su prisutni u eksperimentu. Zanimljivo je kako lako efekti koji se uzimaju kao argument za penetrabilnost vida nestaju kada se sakrije prava svrha eksperimentalne manipulacije (Firestone i Scholl, 2017). S druge strane, ostaje otvoreno pitanje u kojoj mjeri je uopće moguće razdvojiti vidnu percepciju od interpretacija procjena i prosudbi, o čemu postoje podijeljena mišljenja (Durgin, 2017; Schnall, 2017). Nije dovoljno ostati samo na teorijskim raspravama o ovoj distinkciji, nego i u eksperimentima osigurati da subjektivne procjene ispitanika ne budu kontaminirane prosudbama nakon završenog perceptivnog procesa.

Na kraju, možemo zaključiti da za sad nema čvrstih argumenata za penetrabilnost vida budući da se gotovo svi biheviooralni nalazi koji se uzimaju kao podrška ovoj ideji zapravo mogu pripisati ili djelovanju neperceptivnih faktora (prosudba, pamćenje, prepoznavanje i usmjeravanje pažnje) ili nedovoljnoj eksperimentalnoj kontroli, što uključuje kontrolu fizičkih obilježja podražaja, kao i kontrolu ispitanikovih pretpostavki o pravoj hipotezi istraživanja. Stoga ostaje otvoreno istraživačko pitanje može li se kreirati istraživanje u kojem će se nedvosmisleno izolirati izravni utjecaj kognicije i/ili emocija na vid.

LITERATURA

Adelson, E. H., i Bergen, J. R. (1991). The plenoptic function and the elements of early vision. In M. Landy i J. A. Movshon (Eds.), *Computational Models of Visual Processing* (pp. 3–20). MIT Press.

Angelucci, A., Levitt, J. B., i Lund, J. S. (2002). Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1. *Progress in Brain Research, 136*, 373–388. https://doi.org/bj2mv3

Anstis, S. (2002). Was El Greco astigmatic? *Leonardo, 35*, 208–208. https://doi.org/cr6ft6

Baker, L. J., i Levin, D. T. (2016). The face-race lightness illusion is not driven by low-level stimulus properties: An empirical reply to Firestone and Scholl (2014). *Psychonomic Bulletin & Review, 23*(6), 1989–1995. https://doi.org/10.3758/s13423-016-1048-z

Balcetis, E. (2016). Approach and avoidance as organizing structures for motivated distance perception. *Emotion Review, 8*(2), 115–128. https://doi.org/f6t4

Banerjee, P., Chatterjee, P., i Sinha, J. (2012). Is it light or dark? Recalling moral behavior changes perception of brightness. *Psychological Science, 23*, 407–409. https://doi.org/10.1177/0956797611432497

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617–629. https://doi.org/10.1038/nrn1476

Barnes, J., i David, A. S. (2001). Visual hallucinations in Parkinson's disease: A review and phenomenological survey. *Journal of Neurology, Neurosurgery & Psychiatry, 70*, 727–733. https://doi.org/10.1136/jnnp.70.6.727

Barrett, L. F., i Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology, 41*, 167–218. https://doi.org/10.1016/s0065-2601(08)00404-8

Bhalla, M., i Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1076–1096. https://doi.org/10.1037/0096-1523.25.4.1076

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*, 123–152.

Bruner, J. S., i Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42*, 33–44.

Bullier, J. (2001). Feedback connections and conscious vision. *Trends in Cognitive Sciences, 5*, 369–370. https://doi.org/10.1016/S1364-6613(00)01730-7

Callaway, E. M. (2004). Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Networks, 17,* 625–632. https://doi.org/10.1016/j.neunet.2004.04.004

Callaway, E. M. (2005). Structure and function of parallel pathways in the primate early visual system. *Journal of Physiology, 566,* 13–19. https://doi.org/drkq3p

Churchland, P. M. (1988). *Matter and consciousness*. MIT Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 1–73. https://doi.org/f4xkv5

Cole, S., Balcetis, E., i Dunning, D. (2013). Affective signals of threat increase perceived proximity. *Psychological Science, 24*(1), 34–40. https://doi.org/f6t5

Crick, F., i Koch, C. (1998). Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature, 391*(6664), 245–250. https://doi.org/10.1038/34584

Dean, A. M., Oh, J., Thomson, C. J., Norris, C. J., i Durgin, F. H. (2016). Do individual differences and aging effects in the estimation of geographical slant reflect cognitive or perceptual effects? *i-Perception, 7*(4). https://doi.org/10.1177/2041669516658665

Durgin, F. H. (2017). Counterpoint: Distinguishing between perception and judgment of spatial layout. *Perspectives on Psychological Science, 12*(2), 344–346. https://doi.org/10.1177/1745691616677829

Durgin, F. H., Baird, J., Greenburg, M., Russell, R., Shaughnessy, K., i Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review, 16*(5), 964–969. https://doi.org/10.3758/PBR.16.5.964

Durgin, F. H., Klein, B., Spiegel, A., Strawser, C. J., i Williams, M. (2012). The social psychology of perception experiments: Hills, backpacks, glucose and the problem of generalizability. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1582–1595. https://doi.org/10.1037/a0027805

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. MIT Press.

Feldman, F., i Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4*(215), 1–23. https://doi.org/10.3389/fnhum.2010.00215

Firestone, C. (2013). On the origin and status of the "El Greco fallacy". *Perception, 42*(6), 672–674. https://doi.org/10.1068/p7488

Firestone, C., i Scholl, B. J. (2014). "Top-down" effects where none should be found: the El Greco fallacy in perception research. *Psychological Science, 25*(1), 38–46. https://doi.org/10.1177/0956797613485092

Firestone, C., i Scholl, B. J. (2015). Enhanced visual awareness for morality and pajamas? Perception vs. memory in "top-down" effects. *Cognition, 136*, 409–416. https://doi.org/10.1016/j.cognition.2014.10.014

Firestone, C., i Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences, 39*. https://doi.org/10.1017/S0140525X15000965

Firestone, C., i Scholl, B. J. (2017). Seeing and thinking in studies of embodied "perception". *Perspectives on Psychological Science, 12*(2), 341–343. https://doi.org/gh4g

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. MIT Press.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology, 4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gantman, A. P., i Van Bavel, J. J. (2014). The moral pop-out effect: enhanced perceptual awareness of morally relevant stimuli. *Cognition, 132*(1), 22–29. https://doi.org/10.1016/j.cognition.2014.02.007

Gilbert, C. D., i Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience, 14*(5), 350–363. https://doi.org/10.1038/nrn3476

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., i Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review, 106*(4), 795–834. https://doi.org/10.1037/0033-295x.106.4.795

Golby, A. J., Gabrieli, J. D., Chiao, J. Y., i Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience, 4*(8), 845–850. https://doi.org/10.1038/90565

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science, 6*, 298–394.

Goodale, M. A., i Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience, 15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Gregory, R. L. (1970). *The intelligent eye*. Weidenfeld and Nicolson.

Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and Cognition, 8*, 1–44. https://doi.org/10.1006/ccog.1998.0372

Grossberg, S. (2013). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks, 37*, 1–47. https://doi.org/10.1016/j.neunet.2012.09.017

Grossberg, S. (2017). Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Networks, 87*, 38–95. https://doi.org/10.1016/j.neunet.2016.11.003

Helmholtz, H. v. (1867/1924). *Treatise on physiological optics* (J. P. C. Southall, Trans. Vol. III). Dover Press.

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Horr, N. K., Braun, C., i Volz, K. G. (2014). Feeling before knowing why: the role of the orbitofrontal cortex in intuitive judgments – an MEG study. *Cognitive, Affective, & Behavioral Neuroscience, 14*(4), 1271–1285. https://doi.org/10.3758/s13415-014-0286-7

Huang, Y., i Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(5), 580–593. https://doi.org/10.1002/wcs.142

Julesz, B., i Bergen, J. R. (1983). Human factors and behavioral science: Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal, 62*(6), 1619–1645. https://doi.org/f6t6

Klein, O., Doyen, S., Leys, C., Magalhaes de Saldanha da Gama, P. A., Miller, S., Questienne, L., i Cleeremans, A. (2012). Low hopes, high expectations: Expectancy effects and the replicability of behavioral experiments. *Perspectives on Psychological Science, 7*(6), 572–584. https://doi.org/10.1177/1745691612463704

Kok, P., Brouwer, G. J., van Gerven, M. A., i de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience, 33*(41), 16275–16284. https://doi.org/10.1523/jneurosci.0742-13.2013

Kveraga, K., Boshyan, J., Adams, R. B., Jr., Mote, J., Betz, N., Ward, N., Hadjikhani, N., Bar, M., & Barrett, L. F. (2015). If it bleeds, it leads: separating threat from mere negativity. *Social Cognitive and Affective Neuroscience, 10*(1), 28–35. https://doi.org/10.1093/scan/nsu007

Kveraga, K., Boshyan, J., i Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience, 27*(48), 13232–13240. https://doi.org/10.1523/JNEUROSCI.3481-07.2007

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*(1), 12–18. https://doi.org/10.1016/s1364-6613(02)00013-x

Lamme, V. A. F., i Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience, 23*(11), 571–579. https://doi.org/10.1016/s0166-2236(00)01657-x

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27,* 889–935.

Levin, D. T., i Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General, 135*(4), 501–512. https://doi.org/10.1037/0096-3445.135.4.501

Long, G. M., i Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychological Bulletin, 130*(5), 748–768. https://doi.org/10.1037/0033-2909.130.5.748

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology, 6*(4), 547–569. https://doi.org/10.1007/s13164-015-0253-4

Lupyan, G. (2017a). Changing what you see by changing what you know: The role of attention. *Frontiers in Psychology, 8*(553). https://doi.org/10.3389/fpsyg.2017.00553

Lupyan, G. (2017b). How reliable is perception? *Philosophical Topics, 45*(1), 81–106. https://doi.org/10.17605/OSF.IO/R7SJJ

Macknik, S. L., i Martinez-Conde, S. (2009). The role of feedback in visual attention and awareness. In M. S. Gazzaniga (Ed.), *The cognitive neuroscience* (pp. 1165–1179). MIT Press.

Macpherson, F. (2017). The relationship between cognitive penetration and predictive coding. *Consciousness and Cognition, 47,* 6–16. https://doi.org/10.1016/j.concog.2016.04.001

Marr, D. (1982). *Vision*. Freeman.

Marr, D., i Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194*, 283–287.

Martinez-Conde, S., Cudeiro, J., Grieve, K. L., Rodriguez, R., Rivadulla, C., i Acuna, C. (1999). Effects of feedback projections from area 18 layers 2/3 to area 17 layers 2/3 in the cat visual cortex. *Journal of Neurophysiology, 82*(5), 2667–2675.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics, 66,* 241–251.

O'Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., Jr., i Bar, M. (2017). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition, 47,* 63–74. https://doi.org/10.1016/j.concog.2016.05.003

Olofsson, J. K., Nordin, S., Sequeira, H., i Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biological Psychology, 77*(3), 247–265. https://doi.org/10.1016/j.biopsycho.2007.11.006

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Pessoa, L., i Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*(11), 773–783. https://doi.org/10.1038/nrn2920

Peterson, M. A. (1994). Object recognition processes can and do operate before figure–ground organization. *Current Directions in Psychological Science, 3*, 105–111.

Peterson, M. A., i Gibson, B. S. (1994). Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours. *Perception & Psychophysics, 56*(5), 551–564.

Peterson, M. A., Harvey, E. M., i Weidenbacher, H. J. (1991). Shape recognition contributions to figure-ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance, 17*, 1075–1089.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*, 341–365.

Raftopoulos, A. (2001). Is perception informationally encapsulated? The issue of the theory-ladenness of perception. *Cognitive Science*, *25*, 423–451. https://doi.org/b874mv

Raftopoulos, A. (2009). *Cognition and perception: How do psychology and neural science inform philosophy?* MIT Press.

Raftopoulos, A. (2014). The cognitive impenetrability of the content of early vision is a necessary and sufficient condition for purely nonconceptual content. *Philosophical Psychology, 27*(5), 601–620. https://doi.org/10.1080/09515089.2012.729486

Rao, R. P., i Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87. https://doi.org/10.1038/4580

Ratner, K. G., i Amodio, D. M. (2013). Seeing "us vs. them": Minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology, 49*, 298–301. https://doi.org/10.1016/j.jesp.2012.10.017

Riener, C. R., Stefanucci, J. K., Proffitt, D. R., i Clore, G. (2011). An effect of mood on the perception of geographical slant. *Cognition and Emotion, 25*(1), 174–182. https://doi.org/10.1080/02699931003738026

Rock, I. (1983). *The logic of perception*. MIT Press.

Roelfsema, P. R. (2005). Elemental operations in vision. *Trends in Cognitive Sciences, 9*(5), 226–233. https://doi.org/10.1016/j.tics.2005.03.012

Schnall, S. (2017). Social and contextual constraints on embodied perception. *Perspectives on Psychological Science, 12*(2), 325–340. https://doi.org/10.1177/1745691616660199

Schnall, S., Zadra, J. R., i Proffitt, D. R. (2010). Direct evidence for the economy of action: Glucose and the perception of geographical slant. *Perception, 39*(4), 464–482. https://doi.org/10.1068/p6445

Schubotz, R. I. (2015). Prediction and expectation. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference* (Vol. 3, pp. 295–302). Elsevier.

Shaughnessy, J. J., Zechmeister, E. B., i Zechmeister, J. S. (2012). *Research methods in psychology* (9th ed.). McGraw Hill.

Sherman, S. M., i Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 357*(1428), 1695–1708. https://doi.org/10.1098/rstb.2002.1161

Shine, J. M., O'Callaghan, C., Halliday, G. M., i Lewis, S. J. (2014). Tricks of the mind: Visual hallucinations as disorders of attention. *Progress in Neurobiology, 116*, 58–65. https://doi.org/10.1016/j.pneurobio.2014.01.004

Song, H., Vonasch, A. J., Meier, B. P., i Bargh, J. A. (2012). Brighten up: Smiles facilitate perceptual judgement of facial lightness. *Journal of Experimental Social Psychology, 48*, 450–452. https://doi.org/10.1016/j.jesp.2011.10.003

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., i Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron, 49*(6), 905–916. https://doi.org/10.1016/j.neuron.2006.01.021

Treisman, A. M., i Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Ungerleider, L. G., i Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, i R. W. J. Mansfield (Eds.), *Analysis of Visual Behavior*. MIT Press.

Van Bavel, J. J., Packer, D. J., i Cunningham, W. A. (2008). The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. *Psychological Science, 19*(11), 1131–1139. https://doi.org/10.1111/j.1467-9280.2008.02214.x

Vandenbroucke, A. R. E., Fahrenfort, J. J., Meuwese, J. D. I., Scholte, H. S., i Lamme, V. A. F. (2016). Prior knowledge about objects determines neural color representation in human visual cortex. *Cerebral Cortex, 26*(4), 1401–1408. https://doi.org/10.1093/cercor/bhu224

Vecera, S. P., i Farah, M. J. (1997). Is image segmentation a bottom-up or an interactive process? *Perception & Psychophysics, 59*, 1280–1296.

Vetter, P., i Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition, 27*, 62–75. https://doi.org/10.1016/j.concog.2014.04.007

Waters, F., Collerton, D., Ffytche, D. H., Jardri, R., Pins, D., Dudley, R., Blom, J. D., Mosimann, U. P., Eperjesi, F., Ford, S., i Larøi, F. (2014). Visual hallucinations in the psychosis spectrum and comparative information from neurodegenerative disorders and eye disease. *Schizophrenia Bulletin, 40,* S233–S245. https://doi.org/f579xx

Witt, J. K., Proffitt, D. R., i Epstein, W. (2004). Perceiving distance: a role of effort and intent. *Perception, 33*(5), 577–590. https://doi.org/10.1068/p5090

Woods, A. J., Philbeck, J. W., i Danoff, J. V. (2009). The various perceptions of distance: an alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance, 35*(4), 1104–1117. https://doi.org/10.1037/a0013622

APPENDIX D

A Neurodynamic Model of the Interaction Between Color Perception and Color Memory

Marić, M., & Domijan, D. (2020). A neurodynamic model of the interaction between color perception and color memory. *Neural Networks, 129,* 222–248. https://doi.org/10.1016/j.neunet.2020.06.008

ABSTRACT

The memory color effect and Spanish castle illusion have been taken as evidence of the cognitive penetrability of vision. In the same manner, the successful decoding of color-related brain signals in functional neuroimaging studies suggests the retrieval of memory colors associated with a perceived gray object. Here, we offer an alternative account of these findings based on the design principles of adaptive resonance theory (ART). In ART, conscious perception is a consequence of a resonant state. Resonance emerges in a recurrent cortical circuit when a bottom-up spatial pattern agrees with the top-down expectation. When they do not agree, a special control mechanism is activated that resets the network and clears off erroneous expectation, thus allowing the bottom-up activity to always dominate in perception. We developed a color ART circuit and evaluated its behavior in computer simulations. The model helps to explain how traces of erroneous expectations about incoming color are eventually removed from the color perception, although their transient effect may be visible in behavioral responses or in brain imaging. Our results suggest that the color ART circuit, as a predictive computational system, is almost never penetrable, because it is equipped with computational mechanisms designed to constrain the impact of the top-down predictions on ongoing perceptual processing.

Keywords: adaptive resonance theory; cognitive impenetrability of vision; color vision; memory color effect; predictive coding

1. INTRODUCTION

A long-standing debate exists about the question of whether cognitive processes such as thinking, reasoning, prior knowledge, or expectations can directly alter the content of visual perception (Raftopoulos & Zeimbekis, 2015). One perspective is the complete independence of vision from cognition on the grounds of the computational role of vision in delivering an accurate representation of the external environment (Pylyshyn, 1999; Raftopoulos, 2001, 2009, 2014). According to this view, vision is an informationally encapsulated module detached from other cognitive operations. It is essential to our ability to flawlessly navigate through an environment. Therefore, any extraneous influence on the fast, dedicated processes of constructing surface representations in space may be deleterious to our behavioral success. In accord with this perspective, many empirical findings of cognitive penetrability of vision (CPV) may be regarded as instances of insufficient experimental control over relevant factors (Firestone & Scholl, 2017). For example, Firestone and Scholl (2014, 2015b, 2015c) showed that some of the studies that found evidence for CPV actually confused perception with judgment or memory. Other studies failed to provide proper control over low-level stimulus features that offer alternative explanation of the observed findings (Firestone & Scholl, 2015a). Another methodological shortcoming is that many experiments were run with low statistical power. This may lead to a publication bias, that is, to a tendency to overestimate effect sizes and potentially to false positive findings regarding CPV (Francis, 2012, 2019; Francis & Thunell, 2019).

An opposing perspective argues for the complementarity of perception and cognition based on the idea that cognition supplies contextual information that may disambiguate contradictory sensory evidence, or it may fill in missing parts (Goldstone et al., 2015; Lupyan 2012, 2015a, 2017a, 2017b). A large amount of behavioral and brain data has been accumulated over the years suggesting that vision is indeed cognitively penetrable (O'Callaghan et al., 2017; Newen & Vetter, 2017; Vetter & Newen, 2014). Such findings fit well within a predictive coding framework (Clark, 2013; Hohwy, 2013, 2017). According to this view, the brain constantly generates hypotheses about the external world and compares these predictions against available sensory data. Predictive coding is usually cast in terms of Bayesian inference, whereby prior beliefs in the states of the world are expressed as probabilities that are updated in light of new evidence according to Bayes’ formula. The related formulation is a free energy principle, which quantifies the degree of surprise arising from the mismatch between expectations and bottom-up input (Friston, 2010).

At the neural level, feedback projections are a prominent feature of the visual cortex; they are found at all stages in the visual cortical hierarchy (Ahissar & Hochstein, 2004; Gilbert & Li, 2013; Hochstein & Ahissar, 2002). Feedback connections may communicate predictions about what is likely to occur in the environment and where it might occur (Summerfield & de Lange, 2014; Summerfield & Egner, 2009, 2014). If feedback connections share the same features with feedforward connections, and if they have the same effect on the target cortical area, then it is reasonable to conclude that vision as a predictive system is a penetrable system (Lupyan, 2015a). In contrast, there is evidence to suggest that the nature of feedback processing is quite distinct from feedforward processing because feedback modulates neural activity rather than drives it (Macknik & Martinez-Conde, 2009). In addition, feedback pathways may be anatomically segregated from feedforward pathways (Markov et al., 2013, 2014; Markov & Kennedy, 2013) and may utilize different communication channels (Bastos et al., 2015; Michalareas et al., 2016).

In line with previous discussions, Macpherson (2017) has argued that, in principle, it is possible to dissociate predictions from penetrability. However, she has not offered specific examples of how this dissociation might occur in neural processing. In this work, we attempt to fill this gap by demonstrating that a well-developed alternative theoretical framework exists that deals with predictions and top-down expectations in a different way relative to predictive coding models discussed by Clark (2013) and Hohwy (2013, 2017). This is adaptive resonance theory (ART), which was proposed by Grossberg and colleagues and developed over several decades. The ART is a general theory of cortical information processing designed to solve the question of how the brain achieves stable learning and memory in a constantly changing environment (Carpenter & Grossberg, 2003; Grossberg, 2003, 2013). It has been successfully applied in the modeling and understanding of behavioral and neural data about category learning, memory, expectation, perception, attention, and consciousness. Here, we argue that the computational mechanisms of the ART imply that vision is not cognitively penetrable. We attempt to demonstrate how behavioral and functional neuroimaging data, which have been taken as evidence of CPV, can be reinterpreted as a manifestation of the operation of ART processing components. This analysis leads to the opposite conclusion; that is, vision is not cognitively penetrable.

2. BEHAVIORAL AND NEURAL DATA SUPPORTING COGNITIVE PENETRABILITY OF COLOR VISION

A vast amount of empirical studies accumulated over the last decades have reported evidence for CPV (reviewed in Firestone and Scholl, 2017). To make the argument clear, we restrict our attention to the well-studied domain of color perception. First, we review major behavioral and neuroimaging findings supporting CPV. Delk and Fillenbaum (1965) were among the first to report that object recognition systematically biased color perception. They have found that a red–orange heart appears redder than it actually is. In a similar fashion, Goldstone (1995) has found that the categorization of simple objects into letters and numerals biased perception of their color. During training, letters were arbitrarily associated with red hues, and numerals were associated with violet hues. In the test phase, the letter L and the numeral 8 were presented in the same hue, which was midway between red and violet. Results have shown that participants judged L to be redder and numeral 8 to be more violet than they really were, thus pointing to an effect of an abstract stimulus category (letters vs. numerals) on color perception.

When participants were asked to adjust the color of natural or artificial objects to neutral gray, their responses to objects with intrinsic colors were biased towards complementary colors (Hansen et al., 2006; Olkkonen et al., 2008; Witzel, 2016; Witzel et al., 2011). This is known as a memory color effect. It arises because an object with an intrinsic or a diagnostic color creates a weak color impression consistent with the object's typical color retrieved from memory. For example, a banana induces a perception of yellow, and this mnemonic effect biases observers to offset their response toward yellow's complementary color, namely blue. The most compelling evidence of the cognitive penetrability of color perception was recently provided by Lupyan (2015b), who has reported that adapting to objects with intrinsic colors (tomato) creates stronger afterimages (more vivid colors) than adapting to arbitrarily colored objects (car). In addition, stronger afterimages were created by scenes containing intrinsically colored elements (sky) than scenes with arbitrarily colored objects (book). A particularly striking demonstration of this effect occurring in natural images is known as the *Spanish castle illusion*, which is discussed in depth below. It should also be noted that a recent study failed to find evidence for cognitive penetrability of color perception (Valenti & Firestone, 2019). Participants were asked to identify an object with an odd color in a set of three objects. Interestingly, their judgments were not affected by the objects' shape.

There is also neuroimaging evidence of top-down effects on color vision. Taking advantage of the fact that the Greek language has two color terms distinguishing between light and dark blue, Thierry et al. (2009) have found that Greek speakers exhibited greater neural response, as measured by event-related potential (ERP) waveforms, when discriminating between light and dark blue relative to discrimination between light and dark green. The difference was observed in the visual mismatch negativity (vMMN) and in the P1 waveform, suggesting that language labels affect early or pre-attentive color perception. No such difference between blue and green was observed in English speakers. In addition, it was found that the strength of the vMMN signal observed in Greek participants correlated negatively with the length of their stay in the United Kingdom (Athanasopoulos et al., 2010). However, two subsequent studies using artificial color labels have failed to find a difference between the early ERP components within categories (Clifford et al., 2012; He et al., 2014).

Forder et al. (2017) have recently found language-related modulation of P1 but not vMMN. The P1 waveform reflects the activation of more than 30 distinct extra-striate visual areas, and it is likely to include the activation of V4 and the posterior inferotemporal (PIT) cortex (Luck, 2014). However, under closer examination, it appears that the effect observed by Forder et al. (2017) is rather weak, occurring in a narrow temporal window of only 10 ms. Authors have noted that when they increased the temporal window, the effect of language disappeared. It seems that they employed an analysis strategy that increases the chance of false positive findings (Luck & Gaspelin, 2017). It would be more convincing to demonstrate that the same effect appears in different temporal windows, which would imply that the effect does not depend on arbitrary decisions about the location of a measurement window (Bacigalupo & Luck, 2015). Moreover, grand-averaged waveforms in Forder et al. (2017) showed that the wave modulations in experimental conditions occurring before the stimulus presentation were of a similar size to the modulations observed in the post-stimulus period. This points to the presence of a high level of noise in the data, which makes drawing any conclusion unwarranted (Woodman, 2010). Taken together, these considerations suggest that no convincing evidence exists for the language-related modulation of early components of ERP in response to color stimuli. In contrast, there is much more robust evidence of the modulation of late components such as P2 and P3, which are thought to reflect late post-perceptual processes (Clifford et al., 2012; He et al., 2014).

Apart from the ERP evidence of CPV, which is equivocal, two studies using functional magnetic resonance imaging (fMRI) have reported that it is possible to decode brain signals evoked by color-selective cortical areas during the perception of achromatic objects with diagnostic colors, such as a banana or tomato (Vandenbroucke et al., 2016). This suggests that observers really see a typical color (yellow) while they observe a gray banana because object knowledge penetrates the cortical area V4, which is known to be involved in color perception (Brouwer & Heeger, 2009, 2013). On the other hand, Bannert and Bartels (2013) utilized more complex pictures and have found that it is possible to decode color signals in V1 but not in V4. They have argued that V1 reads out expectations about an object's typical color.

The reviewed findings seem to support the conclusion that color vision is cognitively penetrable (Macpherson, 2012). Deroy (2013) has further proposed that cognitive influences on color vision are restricted to prior knowledge about an object. This is mediated by multi-modal representation that integrates an object's color with its shape, texture, and other attributes. In contrast, Zeimbekis (2013) has argued that the memory color effect arises from a judgment bias in a task that does not constrain strategies that participants may employ when they make color adjustments. Based on his analysis of the tasks and stimulus conditions used in reviewed studies, Zeimbekis (2013) has concluded that there is no compelling reason to admit CPV. In this paper, we show that it is possible to observe a memory color effect or a Spanish castle illusion within a computational architecture designed to protect perception from cognitive influences. We suggest that such effects arise from a temporary disruption of ongoing color processing. In the next four sections, we provide an overview of the general computational principles on which ART is based and its specific instantiation designed to model color perception.

3. ADAPTIVE RESONANCE THEORY

ART was proposed as a solution to the problem of instability of learning and memory in a non-stationary environment. The problem concerns how a neural network learns new recognition codes without eroding already established codes. This is known as a stability–plasticity dilemma (Grossberg, 1980, 2013). It has long been realized that learning algorithms, such as error back-propagation, suffer from catastrophic forgetting when faced with new inputs that fell outside of the domain on which the network was initially trained (French, 1999; McCloskey & Cohen, 1989). For example, the network might be trained on a task to discriminate between various types of birds. However, if the task is suddenly switched to distinguishing flowers, the network will eventually forget its acquired knowledge about the birds. Furthermore, error back-propagation typically requires thousands of repetitions of the same patterns to learn the appropriate categories. On the other hand, human learning may occur in a single exposure and may persist throughout life. Interestingly, the same learning problems identified in error-backpropagation persist even in state-of-the-art deep neural networks, and several remedies have been put forward to overcome them (Kirkpatrick et al., 2017; Pfülb et al., 2018; Velez & Clune, 2017; Waldrop, 2019).

According to Grossberg (1980), the solution to the problem of catastrophic forgetting is to endow a neural network with the capability of discriminating between new and old patterns. This is achieved by dividing the network into two subsystems: attentional and orienting. An attentional subsystem is dedicated to processing familiar patterns. It compares sensory (bottom-up) data with learned (top-down) expectations; if the input pattern matches with one of the previously learned codes (categories), then it is recognized as a familiar item, and resonance develops between the sensory pattern and the chosen category. On the other hand, if a sufficiently large mismatch occurs between the input pattern and the learned code, then an orienting subsystem is activated. It produces a reset signal that inhibits the currently active node in the attentional subsystem and initiates a search for another learned code that may match with the input. If no learned code is available that is sufficiently similar to the input pattern, then new neural tissue is committed to learn the presented pattern and to establish a new category. Therefore, the previously learned codes are protected from erosion by novelty detection – those codes will not be erased in the presence of the new patterns; rather, new neural tissue is committed to encode them. Here, we suggest that the same mechanism of novelty detection that is needed to achieve stable learning is also responsible for protecting perception from top-down influences. In other words, the cognitive impenetrability of visual perception is a natural consequence of the brain mechanisms that are designed to prevent interference between new inputs and old memories.

Computational implementation of the ART circuit consists of three hierarchically organized layers labeled as F_0, F_1, and F_2 (Fig. 1). First, the F_0 layer represents a purely sensory response that is not affected by expectations. In some descriptions of the ART circuit, this layer is left out as a less important component. However, its output enters into the computation of a match in the orienting subsystem. Furthermore, it has properties that are relevant for the current discussion. In particular, the F_0 layer should not receive feedback from higher-level centers in the hierarchy. For this reason, we explicitly included it in the description of the ART circuit. Second, the F_1 layer serves as a matching point between bottom-up input arriving from the F_0 layer and top-down signals arriving from the F_2 layer. Third, at the top of the hierarchy, the F_2 layer represents categories or feature groupings. This layer is a winner-takes-all (WTA) network that enhances the activity of the node receiving maximal input and inhibits all other nodes receiving less support. Each F_2 node is specifically tuned to one spatial pattern registered in the F_1 layer.

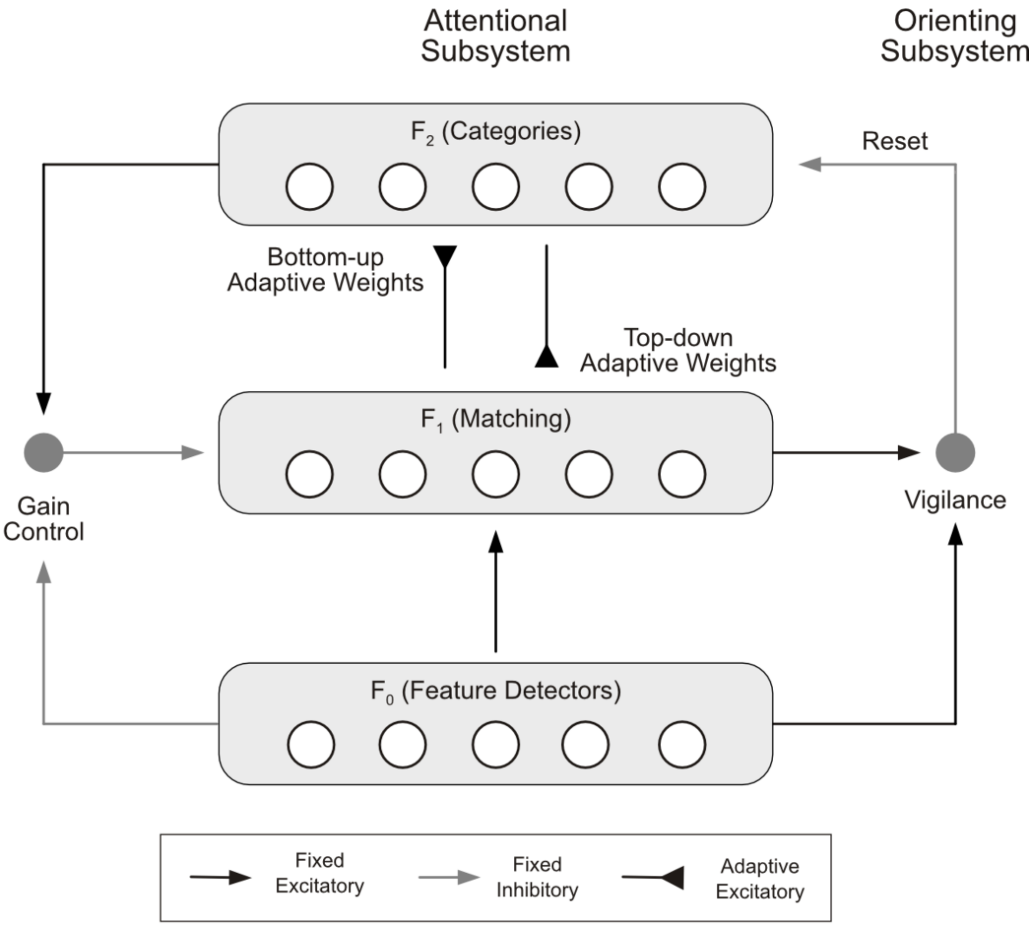


Figure 1. *Canonical ART circuit with parallel attentional and orienting subsystem. The attentional subsystem consists of three neural layers labeled as F0, F1, and F2. In addition, gain control prevents the suprathreshold activation of F1 in the absence of bottom-up input in F0. In the orienting subsystem, an inhibitory reset signal is released toward F2 when a mismatch is detected between the bottom-up input and top-down learned expectation.*

The input pattern is processed in two sweeps through an ART hierarchy. First is a feedforward sweep involving a traversal of activity from F_0 via F_1 to F_2. It ends with the selection of a candidate F_2 node representing the best guess regarding the category to which a given input belongs. The second sweep is a feedback sweep starting from the winning F_2 node that generates feedback activation to F_1. When feedback arrives at the F_1 layer, its nodes compute the intersection between two sources of activation, that is, the activity pattern arriving from F_0 and the learned top-down expectations arriving via feedback projections from F_2. If top-down synaptic weights closely match with the bottom-up pattern, then the total activity in the F_1 layer should be close to the total activity in the F_0 layer. In this case, the orienting subsystem remains silent, and mutual excitation between F_1 and F_2 results with resonance. On the other hand, if top-down weights in the feedback pathway from F_2 to F_1 are dissimilar to the input pattern in F_0, then the total activity in the F_1 layer will be lower than that in F_0. Now, the difference in total activity between F_0 and F_1 is sufficiently large to trigger activation of the orienting subsystem. The orienting subsystem releases the reset signal to the F_2 layer. In this way, the orienting subsystem selectively inhibits the currently active F_2 node and initiates the search for another F_2 node that might better capture the properties of the input pattern. If such an F_2 node is not found among the ones that are already committed to coding different categories, then an uncommitted F_2 node is recruited to learn a new pattern and to form a new category representation (Carpenter & Grossberg, 1987, 2003).

With respect to the discussion about CPV, it should be emphasized that it is possible to excite the F_2 layer even before it receives feedforward input from F_1. Such feedback signals may arise from the processing stages located above F_2 in a hierarchy. One such source of excitation is an inter-ART associative map that connects outputs from different ART circuits. An excited F_2 node leads to the activation of the feedback pathway from F_2 to F_1. However, inhibitory gain control prevents suprathreshold activation of F_1 in the absence of bottom-up input from F_0. It assures that feedback influences on F_1 are only subthreshold or modulatory. Feedback from F_2 prepares or sensitizes the F_1 nodes to respond faster and more vigorously to the bottom-up input from F_0. This preparatory role of feedback is called attentional priming because the F_2 node focuses attention on a specific feature combination that is expected to occur in the upcoming stimulus (Carpenter & Grossberg, 1987, 2003). The gain control mechanism is essential for the proper functioning of the ART circuit because it prevents hallucinations from occurring (Grossberg, 2000). Without gain control, the $F_2 \rightarrow F_1$ feedback pathway may generate suprathreshold activity in F_1 even when there is no bottom-up input.

To understand why the ART circuit generates impenetrable percepts, it should be emphasized that the ART is an attractor network. Temporal evolution of its activity can be described as moving through a state (or phase) space toward a stable equilibrium point or attractor. An attractor represents a content-addressable memory to which the network activity is drawn. An important property of the attractor is that it is a low-energy state of the system. In other words, an attractor does not allow the network activity to leave it once the attractor is reached (Haykin, 2009, Chapter 13). This means that the network activity is resistant to external perturbations. Such network behavior is known as a hysteresis. In the ART circuit, hysteresis is a system-level consequence of the strong excitatory loop formed between F_1 and F_2 (Grossberg, 1980). It is interesting to note that there is evidence of hysteretic effects in visual perception where decisions about an ambiguous, changing stimulus often exhibit the influence of recent visual experiences. For example, hysteresis has been observed in studies of stereopsis (Julesz, 1974), binocular rivalry (Buckthought et al., 2009), motion perception (Williams & Sekuler, 1986), and the identification of line drawings (You et al., 2011), as well as in the dynamic perception of objects and scenes (Poltoratski & Tong, 2014).

Hysteresis is also a reason why a network with strong excitatory loops requires an external reset in the first place (Francis et al., 1994). In the ART circuit, each attractor corresponds to the activation of a single F_2 node. The only way to push the ART circuit from one attractor to the next is by the inhibitory reset signal that arrives from the orienting subsystem. Furthermore, when the ART circuit is already in one of its resonant states, the only way to activate the orienting subsystem and to push the network from the current attractor to the next one is to change the bottom-up input. On the other hand, top-down signals have no access to the orienting subsystem and cannot influence the ART circuit in the same way as bottom-up inputs. This asymmetry between bottom-up and top-down signals in guiding information processing within the ART circuit is essential to understand how cognition influences perception in general (Domijan & Šetić, 2016).

4. ADAPTIVE RESONANCE THEORY, COLOR PERCEPTION AND CONSCIOUSNESS

According to Grossberg (1999, 2017b), *all conscious states are resonant states*. We claim that, as a corollary to this proposition, cognition almost never penetrates conscious visual perception. In other words, cognitive impenetrability is a generic property of the ART circuit. Top-down signals carrying predictions can influence only early stages of processing in the ART circuit before resonance occurs. After resonance is established, no further cognitive influences are possible because the resonant state is an attractor state where top-down expectations agree or match with bottom-up input. This attractor represents the conscious experience of a familiar input pattern. In the ART circuit, resonance is possible only when there is not much prediction error. In contrast, the predictive coding model suggests a lack of activity in the visual cortex in this case because there is nothing to communicate to higher levels in the hierarchy (Clark, 2013; Hohwy, 2013, 2017).

The same reasoning as above can be applied in understanding how a basic visual experience, such as color qualia, emerges from the interactions within the ART circuit. We suggest that the conscious experience of color arises from the resonance established between the color category (hue) encoded in F_2 and the specific pattern of color-opponent activation registered in F_0. The implication of this proposal is that color perception is a discrete event because the ART circuit always chooses one winner among many competitors in the F_2 layer. Only the winning F_2 node is in a position to establish a resonant state. This is consistent with studies demonstrating the quantal (all-or-none) nature of conscious visual experiences (Asplund et al., 2014; Sergent & Dehaene, 2004; Vul et al., 2009).

To make the argument clear, suppose that a red strawberry is presented as an input image. We will create an expectation to see a red color in the F_2 layer because we recognize the shape of the strawberry, since we have already had an experience with red strawberries in the past. Such an expectation matches well with the bottom-up activation of feature detectors dedicated to the red color in the F_1 layer. The match between expectation and sensory evidence leads to a resonant state corresponding to a subjective experience of seeing a red color on the strawberry. On the other hand, if a green strawberry is presented instead, we will not perceive a mixture of red and green, as the predictive coding model would suggest. Rather, the ART circuit will detect a mismatch between expectation (red) and bottom-up input (green). Such a mismatch would activate an orienting subsystem, which delivers a reset signal to the currently active category node in the F_2 layer corresponding to the red color and clears the traces of erroneous expectations. We will consequently perceive the green color, after the reset signal clears the erroneous expectation of a red color and initiates a search to find the most appropriate category representation for the given input. Again, we emphasize that the dynamics of the ART circuit is guided by the bottom-up patterns. The orienting subsystem ensures that a disconfirmed prediction does not take any part in the perceptual experience.

The following question then arises: Why then are there so many behavioral and neuroimaging findings pointing in the opposite direction? We suggest that behavioral and neural data supporting CPV actually capture transients of neural activity occurring during the processing of erroneous expectations. We emphasize that ART mechanisms require some time to complete. This is especially true in the case of a mismatch where the sequence of events needs to occur before the ART circuit settles into an attractor corresponding to the best match between bottom-up and top-down signals – that is, to conscious color perception. These events include computation of a match in the F_1 layer, activation of the orienting subsystem, and inhibition of the F_2 layer. If the ART circuit is probed to respond before completion of its search cycle, then it will respond with the color category that is currently activated in the F_2 layer. In the early stage of processing, when the (mis)match is still computed in the F_1 layer, the activity in the F_2 layer is determined by the top-down expectations. Therefore, the circuit response will reflect expectation rather than perception. However, this does not mean that visual perception is penetrable. Rather, it reveals the operations of a neural circuit that struggles to achieve the best possible consistency with bottom-up input.

The description of a perception as a resonant state aligns with other approaches to visual consciousness that emphasize the role of feedback projections from a higher- to a lower-level area in the hierarchy. However, in contrast to models postulating that feedback from the extra-striate to the striate cortex is a neural correlate of consciousness (Escobar, 2013; Lamme, 2001, 2003), the ART circuit can be situated within the extra-striate cortex itself, as illustrated below. Several authors have pointed out that V1 and feedback to it do not directly contribute to conscious vision (Ffytche & Zeki, 2011; Koch et al., 2016; Leopold, 2012; Silvanto, 2015), although not all authors agree on this point (Lamme, 2018; Pascual-Leone & Walsh, 2001; Tong, 2003). In addition, the ART circuit goes beyond other feedback models because it explicates how top-down feedback signals interact with bottom-up inputs. Establishing a resonant state requires computation of the amount of agreement between bottom-up and top-down stimulation; it is not sufficient to simply blend bottom-up and top-down signals into an indistinguishable mixture.

5. VIGILANCE AND THE DEGREE OF MATCH

The ART circuit allows for flexible control of the degree of match between bottom-up and top-down signals that are needed to achieve resonance. Degree of match is controlled by the vigilance parameter. Vigilance might be conceived as a threshold that controls activation of the orienting subsystem. If vigilance is set to a high value, then the ART circuit will encode specific exemplars, whereas if it is set to a lower value, then it will capture more global, average feature groupings. Low vigilance would lead to a more global representation of a category that captures feature averages, that is, prototypes. Here, we argue that learning basic visual feature such as color requires an ART circuit with vigilance set to a high value. In this way, it is possible to encode and discriminate among a significant number of hues encountered in the environment. In other words, we suggest that each discernible color requires its own separate F_2 node akin to a color grandmother cell (Bowers, 2009, 2017).

It might be argued that the number of F_2 units needed to encode all possible hues far exceeds cortical capacity because it is estimated that a human observer can distinguish among 2.3 million hues. However, when luminance variations are excluded, such an estimate reduces to approximately 26,000 discernible hues (Linhares et al., 2008). As a counterargument, it is possible to envision a kind of combinatorial coding where three separate color ART circuits encode red–green, blue–yellow, and black–white dimensions of the same input. They may operate in parallel as a grandmother cohort that jointly creates an experience of hue (Grossberg, 2017a). For example, in response to the same cone input, one F_2 node should be activated in the red-green circuit to represent the amount of redness or greenness, while one F_2 node should be activated in the blue–yellow circuit to represent the amount of blueness or yellowness, and finally one F_2 node should be activated in the black–white circuit to represent the brightness of the input. In this way, demand on the number of required nodes in the F_2 layer is greatly reduced. It would be sufficient to encode approximately 160 color categories in the red–green and blue–yellow circuits to account for the estimated 26,000 distinct hues. Moreover, 100 additional brightness categories, ranging from pure black via different shades of gray to pure white, might account for a full range of color coding. In total, less than 500 F_2 nodes are sufficient to support fully functional color vision. It is estimated that approximately 10,000 neurons are present within 1 mm^3 of cortical volume, of which three quarters are excitatory (Olman, 2015). Therefore, the complete color F_2 layer (with red–green, blue–yellow, and black–white components) may easily fit within one cortical glob whose size is estimated to be about 1 mm (Conway et al., 2007; Conway & Tsao, 2009). Moreover, there is sufficient neural space within one glob to place multiple complete sets of color ART circuits that cover partially overlapping areas of the central visual field. In this way, it is possible to account for the perception of ensemble colors (Chetverikov et al., 2017; Kuriki, 2004; Maule & Franklin, 2015, 2016).

We speculate that such highly specific learning occurs early in life when basic color categories become firmly established. This does not preclude the possibility of learning new hues later in life when environmental pressures require one to do so. Furthermore, observers may have a different amount of experience with the range of colors and may experience them in different orders. This may account for idiosyncrasies and individual differences in color appearance and color categorization (Emery et al., 2017a, 2017b). Learning in the ART circuit takes place by adjusting the synaptic weights in the bottom-up and top-down pathways in response to continuous-valued, color-opponent inputs. The way in which to set up weights is prescribed by algorithms such as fuzzy ART (Carpenter et al., 1991), Gaussian ART (Williamson, 1996), distributed ART (Carpenter et al., 1998), or some other variants and improvements described in literature (Baraldi & Alpaydin, 2002a, 2002b; Masuyama et al., 2018; Vigdor & Lerner, 2007).

6. MODEL DESCRIPTION

To address the issue of how prior knowledge may affect color perception, we follow the same approach previously employed in simulating the neural substrate of symbol grounding (Domijan & Šetić, 2016). We will assume that separate ART circuits exist that are specialized for the processing of color and form. Parallel processing circuits reflect the division of labor between color-selective and form-selective neurons found in the ventral visual stream (Lennie, 1998). The model components and their interactions are depicted in Figure 2. Mathematical description of the model is provided in the Appendix. The shape ART circuit and inter-ART associative map were not explicitly modeled. They are included in the model description in order to explain how feedback signals to the color ART circuit are generated.

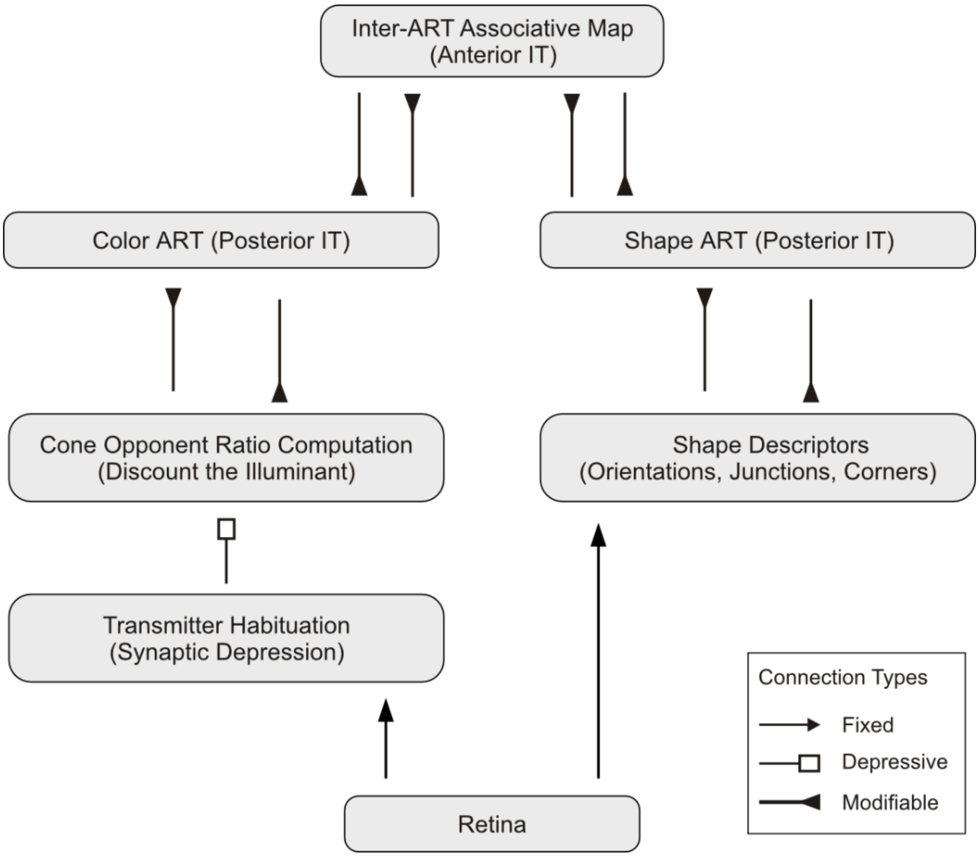


Figure 2. *A model of the interaction between parallel color-selective and shape-selective ART circuits.*

6.1. Shape Adaptive Resonance Theory Circuit

The shape ART circuit learns to recognize distinct shape configurations occurring in the environment. The F_0 and F_1 layers of the shape ART circuit consists of various boundary detectors tuned to properties such as boundary orientations, sizes, terminations, corners, circles, and similar features that are typically found in the form-selective areas of V4 (Pasupathy & Connor, 1999, 2001). The F_2 layer consists of nodes encoding object categories. The shape ART circuit thus transforms the output of boundary-selective nodes into a representation of integrated shape categories. As in the color ART circuit, in the shape ART circuit, the F_2 layer resides in the PIT where neurons selective to complex shape configurations are found (Brincat & Connor, 2004, 2006). Furthermore, input to the shape ART circuit is subject to attentional modulations in order to achieve positionally invariant object recognition.

6.2. Inter-ART Associative Map

Next, we need to explain how object recognition in the shape ART circuit may influence color perception in the color ART circuit. We suggest that the F_2 layers of both ART circuits are mutually connected via the inter-ART associative map (Fig. 2). It establishes the connections between winning nodes in two ART circuits via Hebbian learning and thus multiplexing information about color and form. Such combined selectivity to color and form has been observed in the anterior inferotemporal (AIT) cortex (Chang et al., 2017; Lafer-Sousa & Conway, 2013).

Following O’Callaghan et al. (2017), we suggest that the shape ART module responds faster to the same input pattern than the color ART circuit. This is because shape-selective neurons receive predominantly fast magnocellular input, while color-selective neurons receive predominantly slow parvocellular input. Temporal asymmetry between form and color creates a lag in their encoding in V4. On the one hand, fast processing in the shape pathway agrees with behavioral and neural evidence that object recognition is a rapid process occurring within 150 ms of stimulus presentation (Kirchner & Thorpe, 2006; Thorpe et al., 1996; VanRullen & Thorpe, 2001). On the other hand, surface color may require a slow process of activity spreading or neural filling-in to complete a surface interior (Komatsu, 2006).

Figure 3 illustrates the cortical pathway that enables object knowledge to influence color perception. Fast activation of the shape ART circuit in a response to a familiar shape further activates the inter-ART associative map even before the color ART receives its own bottom-up input. Then, the inter-ART associative map sends feedback signals to the F_2 layer of the color ART circuit. In other words, it reads out the expectation about color that will be seen soon because of prior experience with the presented object. This will occur if a typical or diagnostic object's color exists. In addition, the inter-ART associative map may receive projections from the orbitofrontal cortex, which integrates more general knowledge about an object arriving from different modalities, including vision, audition, and emotion (O'Callaghan et al., 2017).

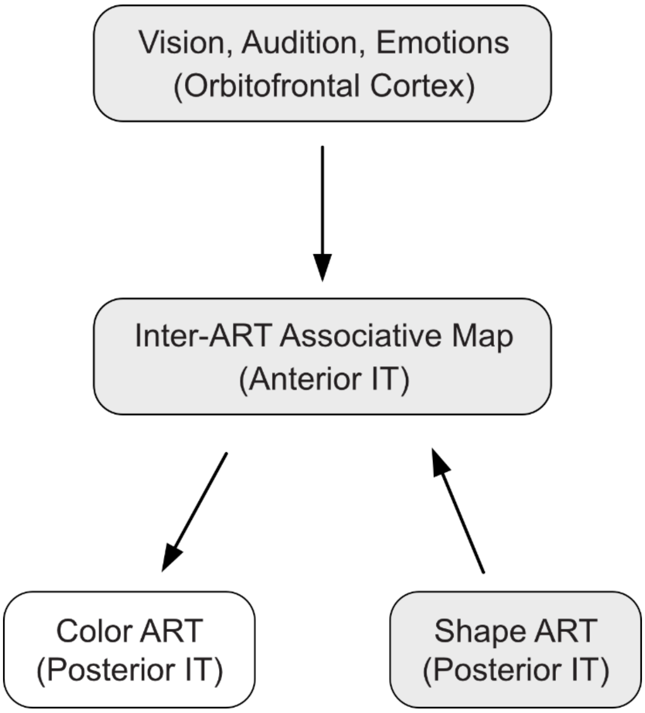


Figure 3. *Cortical pathways that may generate feedback signals to the color ART circuit even before it receives bottom-up input.*

6.3. Color Pre-processing

This model stage encompasses computational mechanisms that take place in front of the color ART circuit (Fig. 4). The goal of this stage is to provide a simplified explanation of how opponent interactions between parallel cone pathways in the retina, LGN and V1 generates the activity of the F_0 layer. To simplify the model, we focus on the red–green opponency only. The F_0 nodes encode the relative activation of the L- and M-cone via the center-surround mechanism. The simplest way in which to achieve this is to compute the difference between them (L-M and M-L). However, we assume that the F_0 layer computes the ratio between the activation of the single cone and the total cone output as in L/(L+M) or M/(L+M). This is achieved by divisive inhibition provided by a common inhibitory interneuron. The advantage of computing a ratio is that it allows F_0 nodes to detach chromatic information from the luminance and thus to discount the incidental variations in illumination and to achieve color constancy (Foster, 2011; Hurlbert, 1996; Smithson, 2005). Discounted luminance information may be encoded in a separate parallel pathway (Hansen & Gegenfurtner, 2009; Lim et al., 2009; Xiao et al., 2003). Another advantage of computing ratio is that it normalizes the activity of F_0 nodes. In addition, early stages of color processing may involve surface filling-in to bind color signals with an object surface (Hong & Tong, 2017; Sasaki & Watanabe, 2004; Seymour et al., 2016). Here, we did not incorporate filling-in because we restrict the model to a single L- and M-cone pathway without tying them to a specific retinal location.

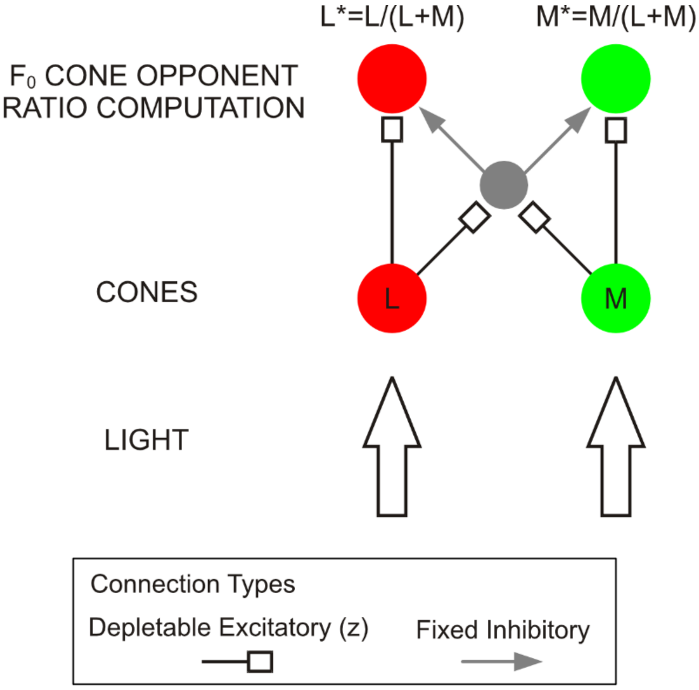


Figure 4. *Gated dipole circuit in cone-opponent pathways. Excitatory connections exist within the same cone pathway, and inhibitory connections exist between pathways. In addition, excitatory and inhibitory connections are endowed with transmitter gates (open rectangles) that adapt and suppress channel output in response to prolonged stimulation.*

To account for the Spanish castle illusion, we need an additional processing component that is capable of generating afterimages. One such component is transmitter habituation or synaptic depression. It refers to a reduction in the amount of available transmitter in response to sustained stimulation of presynaptic sites. Transmitter habituation occurs because transmitter is released at a higher rate than it is possible to synthesize it in presynaptic buttons. A consequence of this process is that a habituated synapse cannot faithfully transmit signal amplitude to the postsynaptic site. When embedded into a competitive circuit with two opponent pathways, which is called gated dipole, transmitter habituation acts as a gate that shifts the competitive balance toward the unhabituated pathway. The gated dipole was designed to explain overshoots and undershoots in neural activity that occur in response to stimulus presentation and withdrawal, respectively (Grossberg, 1980). Also, it helps to explain how afterimages arise in response to prolonged adaptation to color or orientation (Francis, 2010; Francis & Ericson, 2004; Grossberg et al., 2002). In contrast to previous descriptions of the gated dipole circuit, in the current implementation, there is just one inhibitory interneuron receiving excitation from both pathways. In addition, the inhibitory interneuron utilizes divisive instead of subtractive inhibition.

6.4. Color Adaptive Resonance Theory Circuit

The color ART circuit displayed in Figure 5 is based on the real-time implementation of the fuzzy ART algorithm (Carpenter et al., 1991) described in the Appendix. In the color ART circuit, the input (F_0) and the matching layer (F_1) consist of cone-specific nodes. The pre-processing that leads to the activation of F_0 layer is described in the previous section. The F_1 layer computes a fuzzy intersection between the bottom-up input from the F_0 layer and read out of top-down adaptive weights arriving from the F_2 layer. In contrast to F_0 and F_1, the F_2 layer consists of nodes encoding color categories. The F_2 layer is modeled as a WTA network that chooses one candidate node receiving the strongest bottom-up support to represent an input pattern. The F_2 nodes are called hue cells because they selectively respond to the pattern of activation occurring across the F_1 nodes. For example, the F_2 node tuned to the pure red color will respond to the F_1 pattern consisting of a strong activation of the L-cone and a weak activation of the M-cone. In contrast, the F_2 node tuned to the pure green color will respond to the pattern consisting of minimal activation of the L-cone and maximal activation of the M-cone. By the same scheme, all other hue nodes in the range between red and green will be tuned to the particular proportion of the activations of the L- and M- cone.

In this way, the F_2 layer performs the transformation of the linear response of cone-opponent input found in V1 into a nonlinear color-tuned response found in the *globs* regions of the PIT (Bohon et al., 2016; Conway, 2009; Conway et al., 2007; Conway & Tsao, 2009; Zaidi et al., 2014). Globs encode a full hue map because their maximal response shifts systematically across the cortical surface as a function of hue shifts on a color circle. There is also evidence that color-tuned cells exist in V4 and that they cover the complete perceptual color space involving a full set of variations across hue, saturation, and lightness (Li et al., 2014). In addition, functional neuroimaging during execution of color tasks revealed categorical clustering of color representation in V4, as would be expected from the color-tuned cells (Brouwer & Heeger, 2009, 2013).

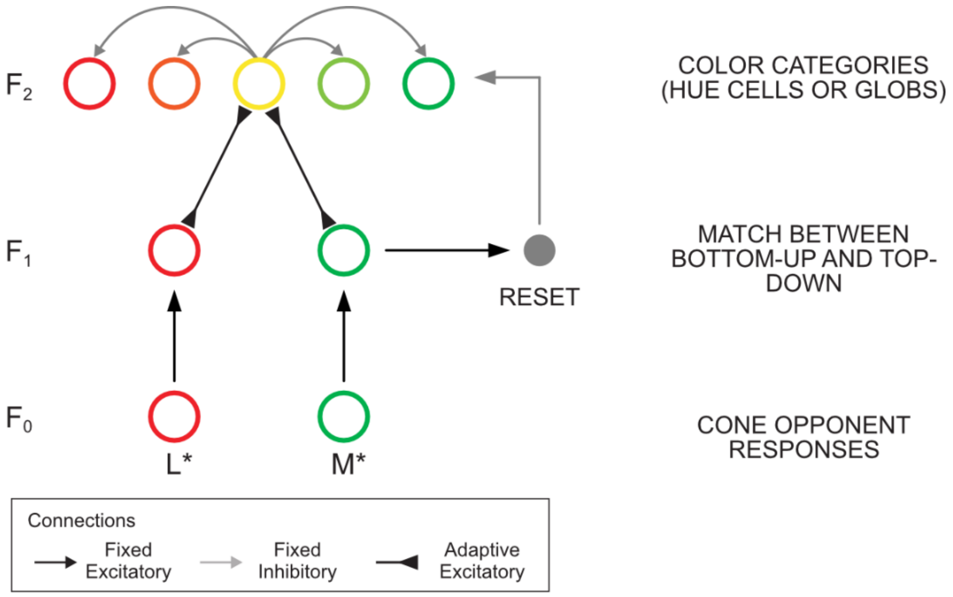


Figure 5. *Color ART circuit, which transforms linear, color-opponent signals registered in F0 and F1 into non-linear, color-tuned responses in F2 where each node encodes a specific hue (denoted by the circle's hue).*

Description of the F_2 layer as a WTA network departs from some versions of the ART circuit, such as distributed ART (Carpenter et al., 1998) or Gaussian ART (Williamson, 1996), where multiple winners are allowed. The distributed ART algorithm was designed to solve the problem of category proliferation and to achieve maximal code compression. On the other hand, the fuzzy ART with high vigilance tends to treat each input pattern as a separate category. In the limit, the fuzzy ART will create as many category representations as there are distinct inputs. Consistent with this observation, many psychological models of category learning assume that each stimulus (exemplar of a category) encountered during learning is stored in long-term memory as a separate item (Estes, 1986, 1994; Lamberts, 2000; Medin & Schaffer, 1978; Nosofsky, 1986, 2011). Exemplar models have been successful in accounting for large datasets of human categorization performance. Therefore, it might be argued that category proliferation observed in the fuzzy ART is not a problem at all. Rather, it suggests that the fuzzy ART might offer an explanation of how exemplars are stored in the brain (Ashby & Rosedahl, 2017). In addition, the WTA network forces the F_2 layer to create a localist representation, that is, the F_2 nodes behave like grandmother cells. Bowers (2009) argued that localist representations provide a better account to many single-unit recording data over distributed representations. He concluded that neuroscience data support embedding of the localist representations in theories of perception and cognition. These observations led us to choose the fuzzy ART model as a psychologically and neurally plausible starting point for the investigation of the interaction between perception and cognition.

6.5. Anatomical Considerations About the Color ART Circuit

We hypothesize that either a complete color ART circuit resides in V4 or it is spanned between V4 and PIT, with V4 encompassing the F_0 and F_1 layers and PIT encompassing the F_2 layer with the representation of color categories. The later interpretation is consistent with the hypothesis that V4 supports figure-ground segmentation by selective extraction of features belonging to a figure while ignoring its background (Papale et al., 2018; Roe et al., 2012). Its position in a visual hierarchy suggests that V4 coordinates signal flow between the early retinotopic maps such as V1–V3 and inferotemporal visual areas involved in recognition and appearance (Winawer & Witthoft, 2015). Irrespective of the exact anatomical location of the color ART circuit, it is not feasible to assume that F_0 and F_1 reside within V1, because such a scheme would require a massive amount of F_2 nodes to separately encode color categories at each retinotopic position in V1. Moreover, each color would need to be exposed in each retinal location to set up the adaptive weights in the feedforward and feedback pathways between F_1 and F_2. To avoid such a combinatorial explosion, we assume that color categories are learned in a spatially invariant representation with converging input from all V1 locations. To avoid mixing of color signals arriving from different locations, spatial attention selectively passes to the color ART circuit the color signal arriving from the target surface and filters out all other surfaces. The way in which this spatial selection might work is explained below in section 8.1. *Adaptive resonance theory and attention*. Such a scheme implies that we can consciously access only a small subset of available colors at any instance of time. Several empirical studies have confirmed that access to feature values is indeed severely limited (Duncan, 1980a, 1980b; Huang & Pashler, 2007; Huang et al., 2007). The idea of a limited conscious experience of color agrees with the hypothesis that consciousness arises in intermediate stages of perceptual processing when we pay attention to a portion of a visual space (Prinz, 2000) and with the claim that visual awareness is rather sparse (Ostergaard, 2018; Ward, 2018).

In contrast to the above conclusion, several studies suggest that observers are able to discern more detailed color information such as the average hue in an array of colored objects (Kuriki, 2004; Maule & Franklin, 2015, 2016). A visual system may even represent the actual shape (uniform or Gaussian) of the distribution of colors in the environment (Chetvertikov et al., 2017). Such an ability to perceive an ensemble of color statistics points to a more detailed phenomenal experience of color than it is possible to verbally report (Block, 2011, 2014). In the ART circuit, it is possible to accommodate the perception of an average hue by noting that spatial attention may be flexibly allocated to a narrow or wide portion of visual space as in a zoom-lens model of spatial attention. When spatial attention is narrowly focused on a small portion of visual space, the F_0 and F_1 layers encode color-opponent responses arriving from a single surface. However, when attention is more widely distributed, the F_0 and F_1 layers may compute averaged color-opponent responses arriving from multiple surfaces. In this case, the F_2 response will reflect an average hue rather than the individual hues of any surface in the ensemble. In support of this explanation, a recent study has found that summary color statistics often go unnoticed (Jackson-Nielsen et al., 2017).

6.6. Color Working Memory Circuit

An important factor that may contribute to the biased reports of color perception in behavioral studies is the transfer of the output from the F_2 layer to working memory (WM) and/or decision-making circuits. Beck and Schneider (2017) have already pointed out that a behavioral response of a participant is not a simple one-to-one mapping of subjective perceptual experience. Rather, it may be distorted on the way from perception to WM and from WM to the response decision. For example, it is conceivable that the output of the ART circuit (activity of the F_2 layer) is further transferred to the color WM in the AIT and to the prefrontal cortex (PFC) where the decision is made regarding the most appropriate response. Here, we assume that the color WM contains the identical hue representation as the one found in the color ART circuit (Fig. 6). We also assume that, like the F_2 layer, the WM circuit is a WTA network; therefore, it picks a hue node with a maximal activity level among the inputs it receives from the F_2. In addition, the input to the WM circuit is passed through a distance-dependent filter in order to account for variability and color confusions observed in the color WM (Allred & Olkkonen, 2015; Bae et al., 2014).

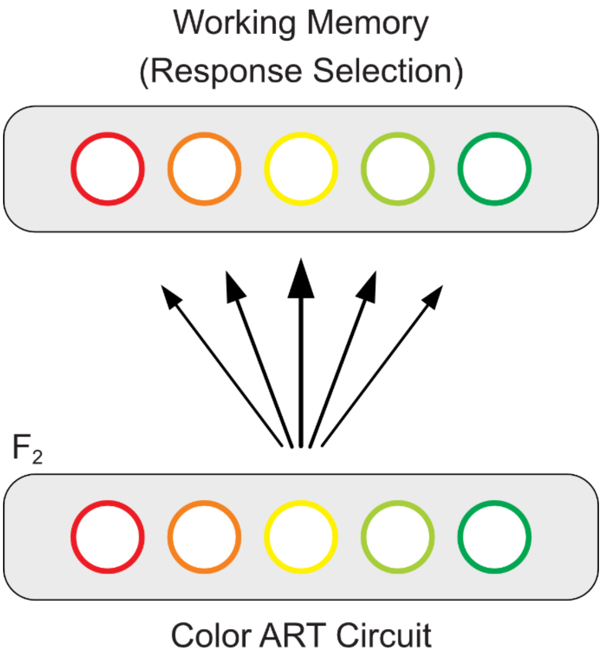


Figure 6. *The transfer of color signals from the F2 layer of the color ART circuit to the visual working memory. The arrows' thickness depicts connection strength.*

Importantly, the transfer from the color ART to the AIT and PFC may commence even before the ART circuit achieves resonance. This claim is consistent with the cascade model of information processing in the brain (McClelland, 1979). It suggests that the flow of activation in cortical layers proceeds in a continuous fashion from earlier to later stages without a clear temporal boundary between them. In the present context, this means that WM and decision-making circuits may pick out traces of top-down expectations even though the ART circuit treats them as erroneous and subsequently removes them from color perception. In other words, the ART circuit has the ability to self-correct its activity in accord with sensory evidence, while the downstream circuits do not have such a capability. As a consequence, observers' responses may be biased because of top-down expectations despite the fact that our color perception is not.

This type of explanation is relevant to all studies where the experimental setup forced participants to compare their actual perception of the currently observed stimulus with the WM representation of the previously attended stimulus (Delk & Fillenbaum, 1965; Goldstone, 1995; Firestone & Scholl, 2014).

7. SIMULATION OF BEHAVIORAL AND NEURAL EVIDENCE OF COGNITIVE PENETRABILITY OF COLOR VISION

7.1. Simulation of Stable Self-organization of Hue Categories

To illustrate that the color ART circuit is able to learn hue categories without catastrophic forgetting, we performed a simulation presented in Figure 7. Input to the network consists of a combination of red and green lights. Light intensity was constrained in the range between 0.0 and 10.0 and varied in steps of 0.5. Each stimulus intensity was presented for one second. The simulation started with the presentation of pure red light (I_L = 10.0, I_M = 0.0) and proceeded with a gradual decrease in the amount of red and simultaneous increase in the amount of green until pure green light (I_L = 0.0, I_M = 10.0) was presented at the end of the learning session. In panels A and B, neural activity is represented as a function of time on a line plot. In panels C–F, neural activity is represented by pixel brightness. Each image has been scaled so that the maximum activity and minimum activity are mapped to black and white, respectively. Intermediate values map linearly onto a gray scale. Node number refers to a spatial location in the network. The size of the F_2 and F_3 layer was set to $N = 21$. Here, we omitted the presentation of activity in the F_0 and F_1 layer. They will be discussed in more detail in the next section.

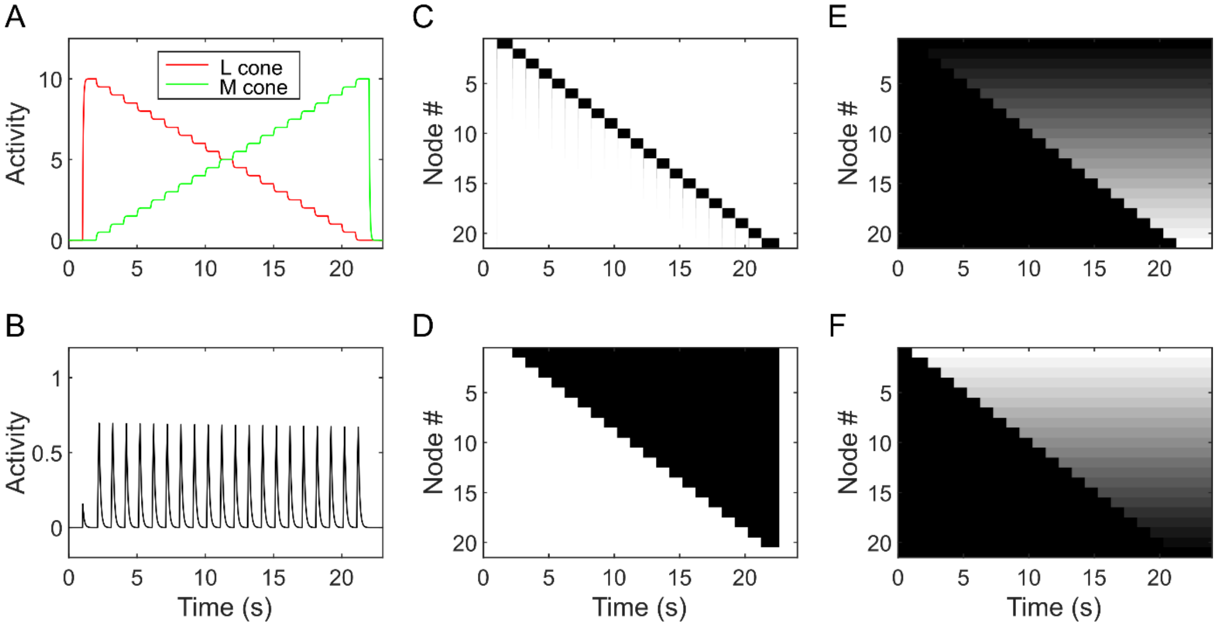


Figure 7. *Unsupervised learning of color categories. (A) The activity of cones, (B) the orienting subsystem, (C) the F2 layer, (D) the F3 layer, and synaptic weights in (E) the L-cone and (F) M-cone bottom-up pathway to the F2 layer are displayed as a function of time.*

As can be seen, each transition in stimulus intensity detected by cones (Fig. 7A) was accompanied by the transient activation of the orienting subsystem (Fig. 7B) that inhibits the currently activated node in the F_2 layer (Fig. 7C). This inhibition enabled activation of the next available F_2 node to encode the new stimulus. The F_2 layer is endowed with an intrinsic signal that varies in intensity as a function of network location. Node 1 received the strongest intrinsic signal, so it wins the competition when the first stimulus was presented. After Node 1 was inhibited, Node 2 won the competition and so on. In this way, the commitment of F_2 nodes to stimuli proceeded in a regular fashion from left to right side of the network. Inhibition to the F_2 layer is delivered by the F_3 layer that detects coincident activation of the node in F_2 layer and the orienting subsystem (Fig. 7D). Importantly, the F_3 nodes remained active after reset in order to prevent subsequent activation of their corresponding F_2 nodes. The F_3 nodes were switched off only at the end of the learning session by the gain control mechanism that multiplicatively gates their self-recurrent collaterals.

Figure 7E and 7F depict learning that occurred in synaptic weights in the bottom-up $F_1 \rightarrow F_2$ pathway originating from the L-cone and M-cone sensitive F_1 node, respectively. The same process occurred in the top-down $F_2 \rightarrow F_1$ pathway so we omit its depiction. Initially, all weights were set to 1. Learning was triggered by the choice of the winning node in the F_2 layer. Synaptic weights of the F_2 wining node were adjusted to match the input amplitudes, that is, the activity pattern in the F_1 layer. For example, Node 1 was the first winner and its weights encoded the pure red light $w_{L1} \approx 1.0$ and $w_{M1} \approx 0.0$. Node 2 was the second winner and its weights encoded the next presented hue $w_{L2} \approx .95$ and $w_{M2} \approx .05$ and so on. After the current winner was inhibited because of the input transition, no further changes in its weights were observed, thus illustrating the stability of learning in the color ART circuit. It should be noted that complement coding in the F_0 layer helps the fuzzy ART to avoid proliferation of F_2 categories and to achieve stable coding of distinct input patterns. Complement coding is a normalization procedure that keeps the magnitude of the input vector (1-norm) constant while preserving the amplitudes of its components (Carpenter et al., 1991).

At the end of the learning session, each F_2 node was tuned to one hue in the range from red to green. The color tuning was spatially ordered as in a rainbow so that we could use the network location as a label for a particular hue. Node 1 is thereby tuned to pure red ($I_L = 10.0$, $I_M = 0.0$), Node 11 to a mid-point between red and green ($I_L = 5.0$, $I_M = 5.0$), and Node 21 to pure green ($I_L = 0.0$, $I_M = 10.0$). All other hues were located between these points. The color tuning established here was used in all simulations reported in subsequent sections.

It might be argued that the order of presentation of stimulus intensities in the previous simulation is not realistic. Early in life, an observer might encounter colors in random order. To address this issue, we ran another learning session with a random order of presentation of the same pairs of light intensities as in the previous simulation (Fig. 8). The F_2 layer was activated in the same orderly fashion from Node 1 to Node 21 because of its intrinsic signal. However, neighboring F_2 nodes encoded widely different hues. Importantly, each F_2 node encoded a distinct hue. When we examined synaptic weights, we found no evidence of merging or omitting color categories. The reason is the fact that the vigilance parameter was set to a high value so that the color ART circuit treated each input pattern as a distinct hue category (i.e., hue exemplar). Therefore, the random order of presentation of light intensities poses no problem for encoding hues.

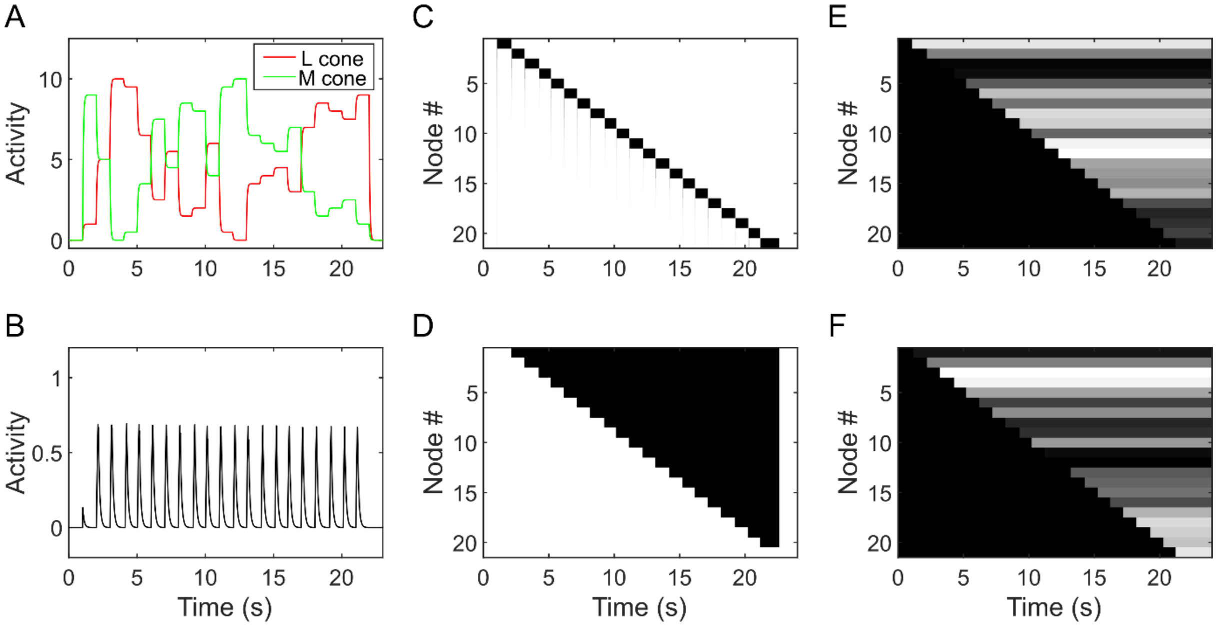


Figure 8. *Unsupervised learning of color categories with a random order of presentation of light intensities. (A) The activity of cones, (B) the orienting subsystem (B), (C) the F2 layer, (D) the F3 layer (D), and synaptic weights in (E) the L-cone and (F) M-cone bottom-up pathway to the F2 layer are displayed as a function of time.*

7.2. Simulation of the Memory Color Effect (Hansen et al., 2006)

Consider an example in which a gray strawberry is presented to the color ART circuit. The bottom-up input was set at $I_L = 5.0$, $I_M = 5.0$ from $t = 200$ ms until the end of simulation. A read-out of color expectation from the inter-ART map creates another input to the F_2 layer.

Here, we assume that a single node in the inter-ART map is connected to multiple nodes in the color ART circuit. The reason for this assumption is that we can observe the same object with different colors in different occasions. Therefore, it is likely that we can store in long-term memory different hues associated with the same object. Moreover, it is possible that object recognition further accesses linguistic knowledge that can also influence the F_2 layer. In this example, recognition of the strawberry may activate the verbal label red that can further activate a range of F_2 nodes associated with this label (Lupyan, 2012).

Figure 9 illustrates two hypothetical distributions of feedback signals, labeled as narrow and wide, whose impact on the activity of F_2 nodes were examined in subsequent simulations. The peak of narrow feedback is centered on the Node 3 near the node encoding pure red (Node 1). Impact on the neighboring nodes falls off rapidly in a distance-dependent manner. Such a narrow distribution corresponds to a situation where color is highly diagnostic, that is, the observer had a narrow range of color experiences associated with an object. In contrast, the peak of wide feedback is centered on Node 7, which is closer to the neutral point (Node 11). Wide feedback falls off more slowly as a function of distance from the peak. This corresponds to a situation where color is less diagnostic because the observer might have more varied color experiences associated with an object. Still, even wide feedback may exert some influence on the F_2 layer and thus bias color perception. In addition, we examined how the systematic variation of the feedback strength influences the activity of the F_2 layer. To this end, we varied the feedback gain factor ϕ that multiplicatively scales the impact of feedback distribution. The feedback gain factor assumed values of 0.1, 0.3, and 0.5. Feedback distributions depicted in Figure 9 induce bias toward red hues. In a subsequently described simulation of the Spanish castle illusion we also employed feedback with bias to green hues. Green-biased feedback distributions were created as mirror images of the red-biased feedback.

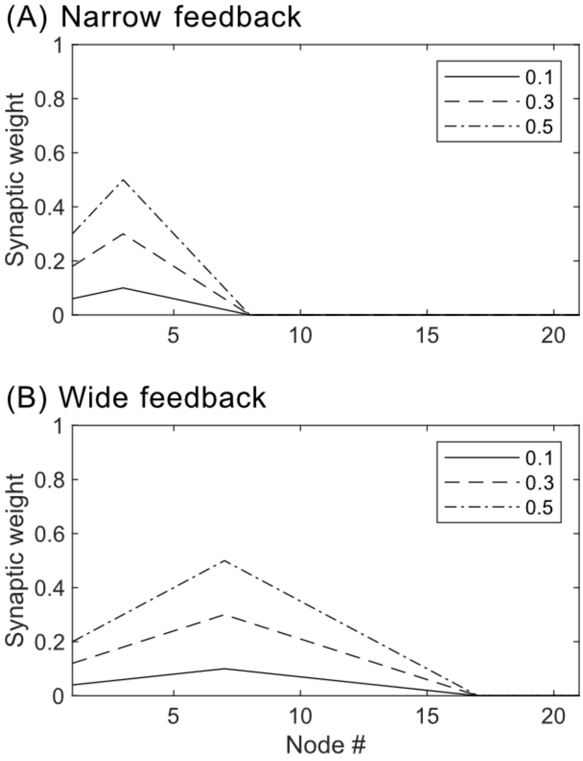


Figure 9. *(A) The narrow and (B) the wide distribution of feedback signals from the inter-ART associative map to the F2 layer were considered in simulations with three levels of the feedback gain factor.*

Figure 10 illustrates the activity of opponent L- and M-cone pathways in response to the presentation of gray color. The activity of cones simply tracked respective input amplitudes (Fig. 10A). Here, we used a convention to slightly offset activity of the M-cone in order to make it visible since its activity was identical to the L-cone. The same convention was used in plots of the transmitter gates and the activity of F_0 and F_1 nodes. In this simulation, transmitter habituation was switched off so the amount of available transmitter was kept constant at the initial value of 1 for the total duration of simulation (Fig. 10B). Consequently, transmitter gates faithfully carried over input amplitudes from cones to the F_0 nodes. The F_0 nodes normalize cone output by divisive inhibition (Fig. 10C). After an initial burst of activity due to disynaptic inhibition, both F_0 nodes settled to 0.5 illustrating their ability to represent relative contributions of the L- and M-cone output.

The F_1 nodes tracked the activity of F_0 nodes except during a short period of time when top-down signals from the F_2 node reduced the activity of F_1 node in the M-cone pathway (Fig. 10D). The reason for this activity reduction was the activation of the F_2 node at location 3 (Fig. 10E). This node was a winner of the competition due to the presence of narrow feedback with a gain factor of 0.1. Here, we assume that the arrival of bottom-up input was delayed relative to the arrival of feedback signals to the F_2 layer. Feedback was thus delivered to the F_2 layer from t = 200 ms until the end of simulation. For this reason, the winning node in the F_2 layer was not a node encoding hue of the bottom-up input, that is, Node 11; instead, the winner was shifted in the direction of the feedback – in this case, Node 3. When the winning F_2 was chosen, it had sent top-down signals to F_1, where a match (i.e., fuzzy intersection) between the bottom-up activity and the top-down weight vector was computed.

Node 3 is tuned to a hue consisting of strong L-cone (I_L = 9.0) and weak M-cone (I_M = 1.0) activation. Such color tuning is reflected in its top-down weights to F_1 where $w_{3L} \approx .90$ and $w_{3M} \approx .10$. Since the F_1 nodes computed a fuzzy intersection between the bottom-up input and read out of the top-down weights, it is clear that the activity reduction in the M-cone pathway arises from the weak top-down signal. On the other hand, no such reduction was observed in the F_1 node of the L-cone pathway because its activity was already the fuzzy intersection between the bottom-up signal of 0.5 and top-down signal of 0.9. Interestingly, during this time period, Node 3 remained fully activated despite reduction in its bottom-up input. Such sustained activation or hysteresis is a consequence of a strong self-excitation.

It might be argued that the activity reduction observed in the F_1 node of the M-cone pathway actually supports CPV and provides a mechanistic account for it. However, this is only the first part of the full processing cycle in the ART circuit. The orienting subsystem continuously monitors for a possible mismatch between the total activity in the F_0 and F_1 (Fig. 10F). Reduction of the total F_1 activity triggered the transient activation of the orienting subsystem that inhibited the currently winning F_2 node. Thereby, the orienting subsystem enabled a search for a new F_2 category node that might achieve a better match with the F_0 pattern. In the next attempt to find an appropriate hue category, bottom-up activity outweighed the influence of feedback and the F_2 layer selected the hue node most similar to the bottom-up input, that is, Node 11. A newly activated F_2 node generated its own top-down signals toward F_1. The top-down weights of Node 11 are $w_{11L} \approx .50$ and $w_{11M} \approx .50$. Thus, the fuzzy intersection allowed the F_1 node of the M-cone pathway to recover from the previous erroneous feedback and to reinstate its bottom-up activity level of 0.5. In this case, the F_0 and F_1 activity patterns matched well so there was no further activation of the orienting subsystem. As a result, the network continued to support the activation of Node 11 in the F_2 layer to the end of simulation. In other words, the network established a resonant state supporting the observer's unbiased conscious perception of the gray color in the strawberry. By unbiased perception, we mean color perception that was not influenced by memory retrieval or expectation. Importantly, feedback was continuously present from t = 200 ms until the end of the simulation.

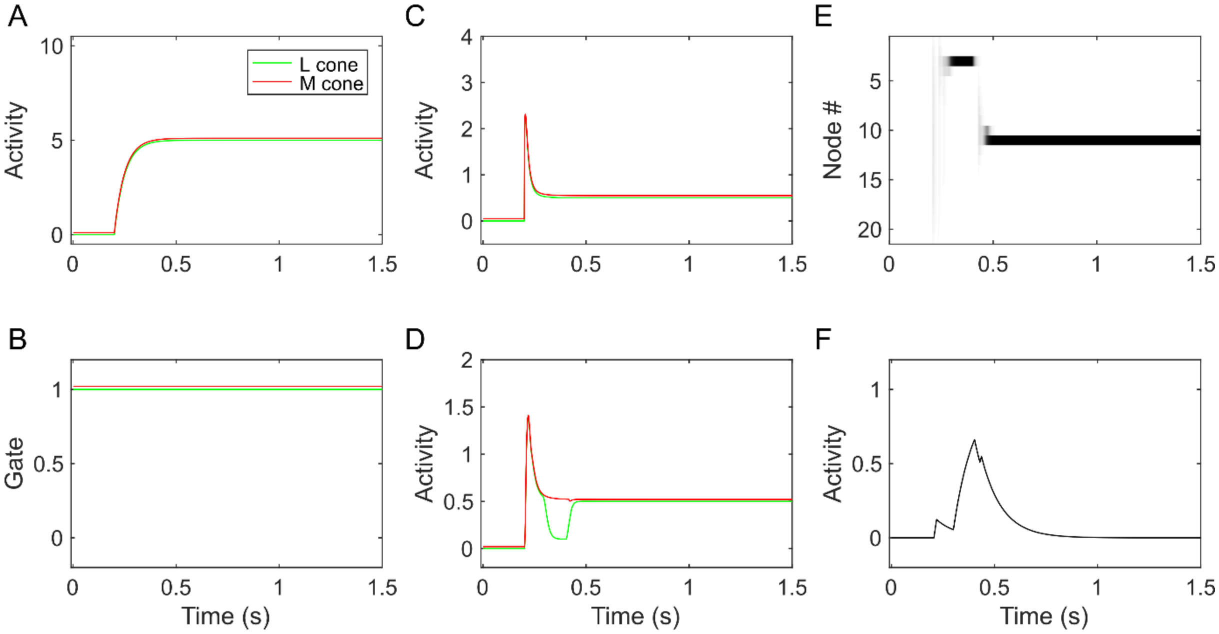


Figure 10. *Simulation of the memory color effect. (A) The activity of cones, (B) transmitter gates, (C) the F0 layer, (D) the F1 layer, (E) the F2 layer, and (F) the r node of the orienting subsystem are displayed as a function of time.*

Figure 11 illustrates how the F_2 layer handles the influence of narrow and wide feedback at three levels of the feedback gain factor 0.1, 0.3, and 0.5. As in the previous simulation, the bottom-up input was I_L = 5.0 and I_M = 5.0 from t = 200 ms until the end of simulation. When the gain factor of narrow feedback was weak, that is 0.1, the F_2 node initially chose Node 4 as a winner (Fig. 11A top). This is close to the peak of the feedback distribution centered on Node 3. However, the feedback was weak enough to allow bottom-up input to slightly bias the competitive balance toward the center of the network. When the gain factor was increased to 0.3, the F_2 layer initially chose Node 3 because narrow feedback was strong enough to overweight the contribution of the bottom-up input (Fig. 11B top). Next, when the feedback strength was further increased to 0.5, the F_2 node made several attempts to find a resonant state starting from Node 3 (Fig. 11C top). In this condition, narrow feedback was strong enough to almost completely overshadow the bottom-up input. In general, as the strength of feedback signal increased, the F_2 layer spent more time in searching for the resonant state. In other words, it spent more time in a biased or penetrable state. A further increase in the feedback strength would result in an exhaustive search of all nodes positioned at locations 1–10. Such a timeconsuming process is clearly detrimental for an individual coping with a fast-changing environment. It is thus possible that neural tissue protects itself from long, unnecessary memory searches by imposing an upper limit on the strength of feedback connections. A comparison of Figures 11B and 11C suggests that ϕ = 0.3 is a reasonable upper limit, so we used this setting in all subsequent simulations.

Wide feedback with the gain factor of 0.1 produced a similar pattern of activity as narrow feedback did (Fig. 11A bottom). When the gain factor of wide feedback was increased to 0.3, the F_2 layer underwent a series of transitions starting with the selection of Node 7, jumped to Node 9 in the interval between 400 ms and 600 ms, then moved to Node 10 between 600 ms and 800 ms, and eventually reached Node 11 after approximately 800 ms (Fig. 11B bottom). When the gain factor was set to 0.5, the F_2 layer also visited Node 8, thus making an even longer search for the match with the bottom-up input (Fig. 11C bottom). An important observation with respect to CPV is that in each instance examined, the F_2 layer eventually reached Node 11, which corresponds to the exact mid-point between red and green. However, the time needed to select this node varied as a function of the strength and shape of the distribution of feedback signals. In general, narrow feedback created stronger but short-lived hue bias, and wide feedback created weaker but longer-lasting hue bias.

We would like to emphasize that the parameters of the feedback distribution were chosen arbitrarily from a range of possible values. However, they should be treated as the model’s free parameters that can be adjusted to fit the empirical data. As the simulation in Figure 11 illustrates, moving the peak of the feedback distribution toward Node 1 (or Node 21) makes the memory color effect stronger. Conversely, moving the peak closer to Node 11 makes the effect weaker. In addition, widening the spread of feedback distribution as well as increasing the feedback gain factor enabled the network to spend more time in a biased state, thus increasing the chance to detect the memory color effect.

Further empirical work is needed to quantify the strength and shape of feedback influences and to constrain the choice of the model’s parameters. In this regard, an important practical difficulty is how to control for an observer’s prior experience with various color-object pairs. It is reasonable to assume that such experiences are subject to great inter- and intra-individual variability. One solution to this problem would be to use artificial stimuli instead of natural objects (Goldstone, 1995). For example, in a training session, the experimenter may control for the frequency of occurrence of all color-shape pairs prior to the test session where the size of the memory color effect is assessed. Such a study may shed light on the relationship between the extent of prior experiences with object-color pairs and the strength of the memory color effect they produce (the same suggestion extends to the Spanish castle illusion discussed in the next section).

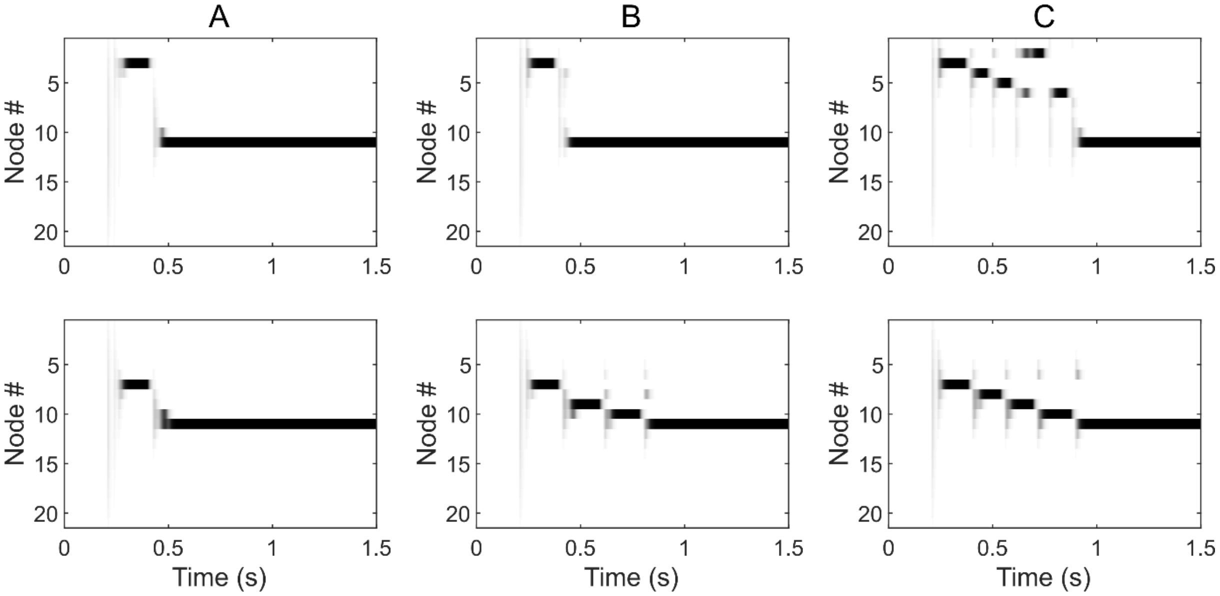


Figure 11. *How the strength and duration of the memory color effect vary as functions of the type and strength of feedback signals that were applied. The activity of the F2 layer is illustrated in response to a joint presentation of a gray color and (top row) the red-biased narrow or (bottom row) red-biased wide feedback. The feedback gain factor ϕ was set to (A) 0.1, (B) 0.3, or (C) 0.5.*

Figure 12 illustrates how the timing of arrival of feedback signals relative to the bottom-up input influenced the activity of the F_2 layer. As in previous simulations, the bottom-up input was I_L = 5.0 and I_M = 5.0 from t = 200 ms until the end of simulation. Moreover, the feedback gain factor was fixed at 0.3 in both narrow and wide feedback. Feedback was turned on at t = 250 ms, t = 275 ms, or t = 300 ms, thus creating feedback delays of 50 ms, 75 ms, or 100 ms, respectively. When feedback was delayed for 50 ms, the F_2 layer was biased toward red hues to the same degree as observed when there was no delay in both narrow (Fig. 12A top) and wide (Fig. 12A bottom) feedback. However, when the feedback delay was 75 ms, narrow feedback was unable to penetrate the activity of the F_2 layer (Fig. 12B top). In the case of wide feedback, there was a weak bias toward red, as the initial winner was Node 10 (Fig. 12B bottom). Later on, this was corrected by the reset signal, and the F_2 layer chose Node 11. When the feedback delay was 100 ms, neither narrow (Fig. 12C top) nor wide (Fig. 12C bottom) feedback could penetrate the F_2 layer anymore. In this condition, the F_2 layer immediately chose Node 11 as a winner. This simulation indicates that there is a narrow temporal window during which feedback signals are able to penetrate the F_2 layer and to bias color perception toward diagnostic hues.

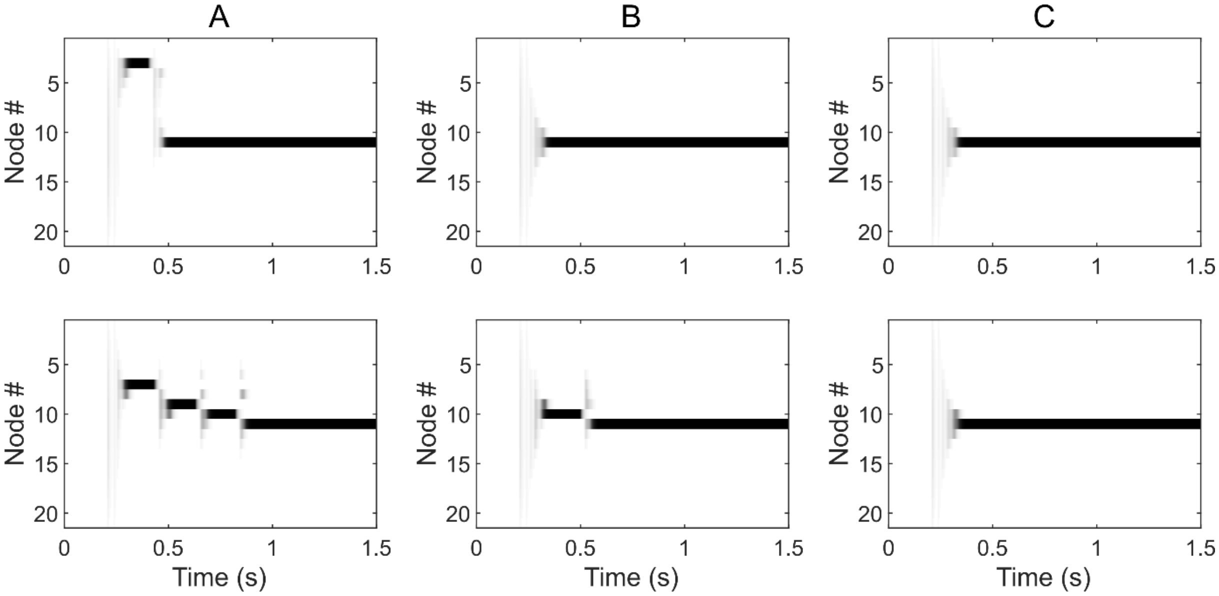


Figure 12. *The activity of the F_2 layer is shown in response to a joint presentation of a gray color and (top row) the red-biased narrow or (bottom row) red-biased wide feedback, with a feedback delay of (A) 50 ms, (B) 75 ms, or (C) 100 ms.*

Figure 13 illustrates how the speed of neural activation (a node's time constant τ_x) influenced the penetrability of the F_2 layer. We used the same settings for the bottom-up input and for narrow and wide feedback as in the previous simulation. In addition, we varied the time constant of the F_2 node between 100 ms, 50 ms, and 25 ms to make it slower relative to the F_1 layer and other model components. When the F_2 layer was made very slow with τ_x = 100 ms, the memory color effect completely disappeared in the condition of narrow feedback (Fig. 13A top). On the other hand, wide feedback produced a weak but long lasting memory color effect (Fig. 13A bottom). The F_2 layer selected Node 10 as a winner and this node remained active almost to the end of simulation. At around 1,400 ms, Node 10 was inhibited and eventually replaced by Node 11 (not shown). When the speed of neural activity was set at τ_x = 50 ms, narrow feedback still did not show a memory color effect while wide feedback exhibited a stronger but short lasting effect (Fig. 13B top). Under wide feedback, the F_2 layer selected Node 8 but it was inhibited and replaced by Node 11 at around 1,000 ms (Fig. 13B bottom). Finally, when the time constant was set at τ_x = 25 ms, both narrow (Fig. 13C top) and wide (Fig. 13C bottom) feedback produced a memory color effect that was comparable in magnitude and duration to the effect observed with the default setting of time constant $\tau_x = 10$ ms.

As can be seen from the previous simulations, achieving unbiased color perception requires some time to clear the traces of top-down influences. The amount of time required to reach resonance depends on several factors. First, it depends on the strength of feedback signals and the shape of their distribution. Second, it depends on the relative timing of feedback signals and bottom-up input. Finally, the speed of neural integration within nodes in the F_1 and F_2 layers (their time constants) may also contribute to the total amount of time the network will spend in the hysteretic state that is not supported by the bottom-up input. If the observer chooses to respond prior to the occurrence of the reset signal, then his or her report will reflect bias induced by the top-down expectation. However, if the observer waits for a while until the dynamics of the ART circuit settle on the unbiased, long-lasting resonant state, then no memory color effect would occur as observed by Valenti and Firestone (2019). Consistent with this observation, Olkkonen et al. (2008) have found that the strength of the memory color effect negatively correlates with response times: the longer participants wait before issuing a response, the weaker the effect becomes. On the other hand, Lupyan (2015b) has not found such a correlation, although it should be noted that, in the instructions to participants, he emphasized that they should respond as fast as possible. This pressure to respond fast, we argue, creates a bias to report color that reflects top-down influences because the color ART circuit did not settle on its final color category yet.

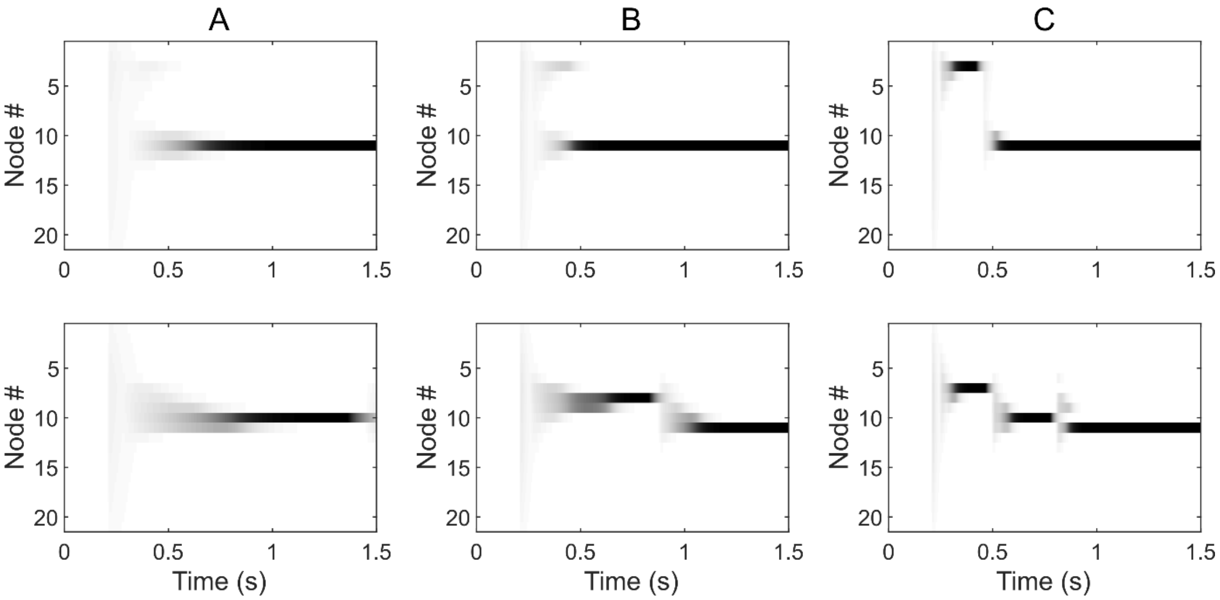


Figure 13. *The activity of the F2 layer is depicted in response to a joint presentation of a gray color and (top row) the red-biased narrow or (bottom row) red-biased wide feedback, with the integration time constant of the F2 nodes set to (A) 100 ms, (B) 50 ms, or (C) 25 ms.*

It should be noted that in all simulations reported in this paper, the vigilance parameter was set to a high value. It might be argued that this is not a realistic scenario and that there is a possibility for lapses in vigilance that may lead to acceptance of biased hues in the color ART circuit. This is a potential avenue for allowing cognition to penetrate color perception. However, we would like to emphasize that any such lapses of vigilance are transient in nature. Even if the ART circuit spends some time in a state of low vigilance where resonance is established with a biased hue category, it will eventually recover from such condition and reinstate its sharp distinction between hue categories. We consider variations in vigilance as just another factor that contribute to a great variability in the strength of a memory color effect as observed experimentally.

Figure 14 illustrates the behavior of the F_2 layer under different choices of vigilance parameter ρ. The bottom-up input as well as narrow and wide feedback were the same as in previous simulations. Vigilance was set at 0.5, 0.7, or 0.9 in the interval [0 ms, 800 ms]. A low level of vigilance such as 0.5 enabled both narrow (Fig. 14A top) and wide (Fig. 14A bottom) feedback to penetrate the F_2 layer and to generate a resonant state with a feedback-biased choice of hue. When vigilance was increased to 0.7, narrow feedback produced a sufficiently large mismatch that was detected by the F_2 (Fig. 14B top). Subsequently, Node 11 was chosen that corresponds to veridical perception of gray. On the other hand, wide feedback passed the vigilance test and remained in a biased state with the choice of Node 7 as a winner (Fig. 14B bottom). When vigilance was further increased to 0.9, even wider feedback produced reset of the F_2 layer (Fig. 14C bottom). However, after reset, Node 9 was chosen as a winner suggesting that the network remained in a biased state. Importantly, after vigilance returned to its default value of 0.96 at t > 800 ms, in all conditions examined, the F_2 layer reacted with the inhibition of the current winner and selection of Node 11 unless Node 11 was already selected before.

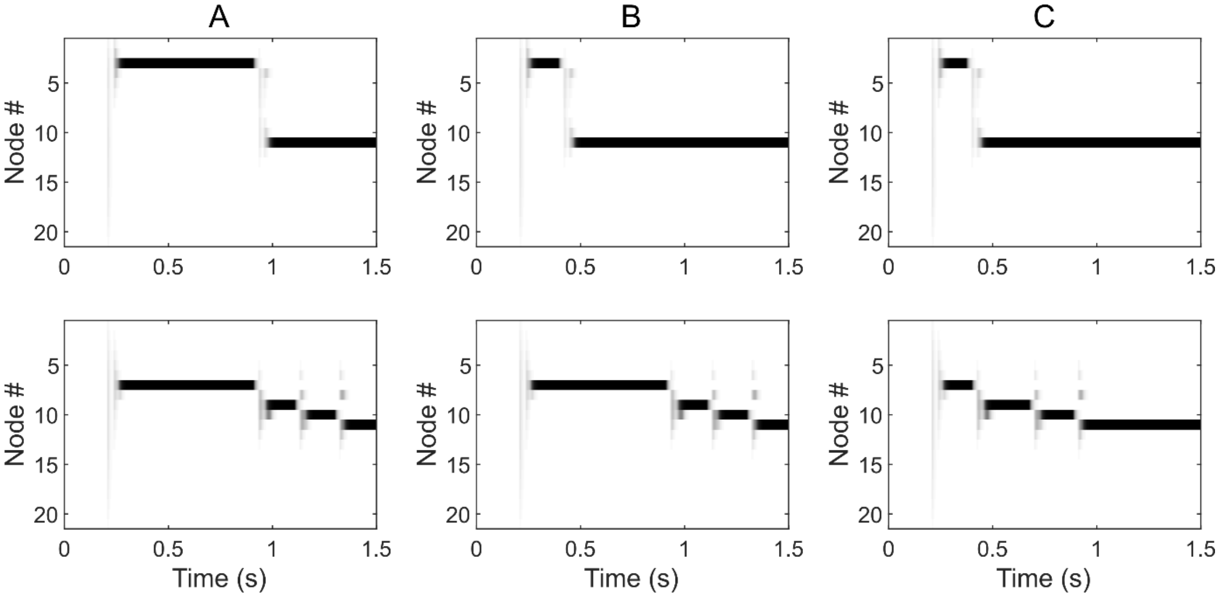


Figure 14. *The activity of the F2 layer is portrayed in response to a joint presentation of a gray color and (top row) the red-biased narrow or (bottom row) red-biased wide feedback, with a vigilance parameter ρ set to (A) 0.5, (B) 0.7, or (C) 0.9 from the start of the simulation up to 800 ms. After 800 ms, vigilance was set to a default value of 0.96 until the end of the simulation.*

7.3. Simulation of Spanish Castle Illusion (Lupyan, 2015b)

Figure 15 illustrates how color pre-processing generates a typical afterimage without the presence of feedback signals. First, we let the network adapt to the red color followed by the presentation of gray color. Thus, the bottom-up input was set to I_L = 10.0 and I_M = 0.0 in the interval between t = 100 ms and t = 600 ms and then to I_L = 5.0 and I_M = 5.0 until the end of simulation. During the maintenance of an elevated activity level in the L-cone pathway (Fig. 15A), its neurotransmitter gate decayed and became less effective because of the exhaustion of its presynaptic buttons (Fig. 15B). In other words, prolonged exposure to the same stimulus creates a situation where presynaptic neural activity places more demands on the transmitter release than the presynaptic buttons can synthesize.

In the F_0 layer, there was an initial overshoot in the L-cone pathway in response to the presentation of red color (Fig. 15C). It was followed by a small dip in its activity because of transmitter depletion. However, it was counteracted by the similar depletion in the inhibitory connection to F_0. Here, divisive inhibition enables the F_0 node to remain sensitive to its input despite the fact that the transmitter gate almost approached zero.

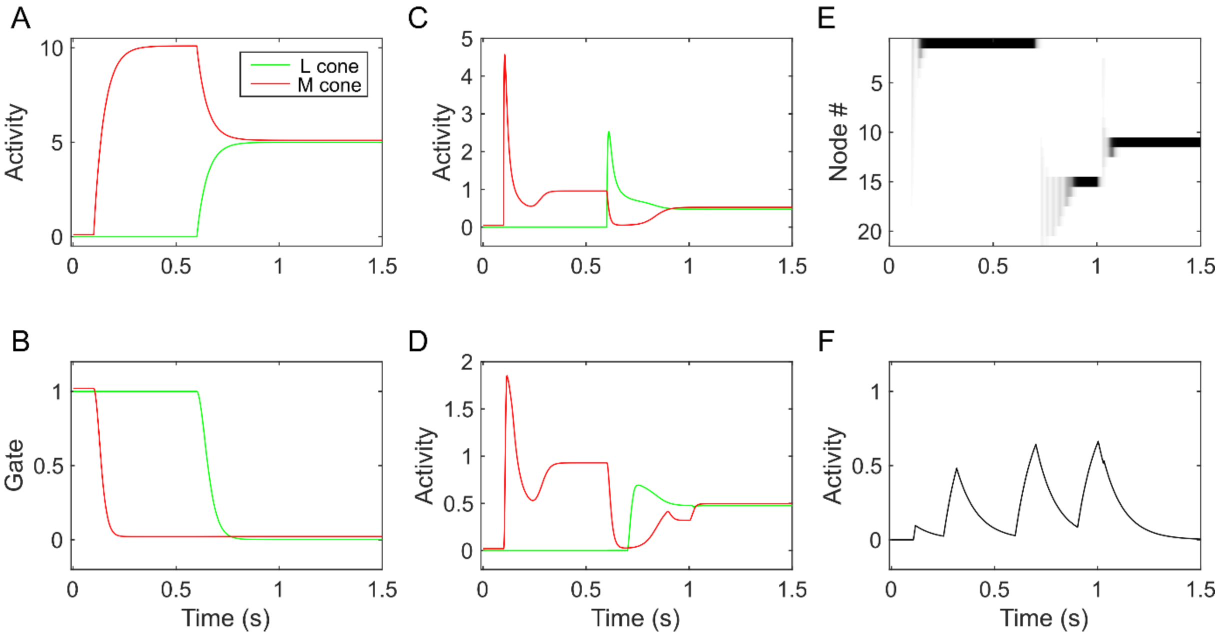


Figure 15. *Simulation of an afterimage without feedback. (A) The activity of cones, (B) transmitter gates, (C) the F0 layer, (D) the F1 layer, (E) the F2 layer, and (F) the r node of the orienting subsystem are displayed as a function of time.*

After the network was adapted to red, the subsequent presentation of a mid-gray stimulus activated both cone pathways to the same degree. However, the response of the F_0 node in the M-cone pathway was stronger than that of the L-cone because of the imbalance in the amount of neurotransmitter available in the two pathways. As a consequence of this imbalance, there was an activity overshoot in the M-cone pathway. The temporal evolution of activity of the F_1 nodes follows a similar pattern as observed in respective F_0 nodes (Fig. 15D). Their activity level was generally lower relative to F_0 nodes because they compute a fuzzy intersection. The activity of the F_2 layer (Fig. 15E) and the orienting subsystem (Fig. 15F) tracked changes that occurred in the F_1 layer.

At the beginning, the F_2 layer selected Node 1 as a winner, indicating perception of pure red. When the input color switched to gray, activity overshoot in the F_1 node of the M-cone pathway triggered the activation of the orienting subsystem and forced the selection of Node

15 in the F_2 layer. This node is tuned to input $I_L = 3$ and $I_M = 7$ that, in normal circumstances, corresponds to perception of slightly greenish hue. Node 15 is far from pure green (Node 21) but it still represents a bias toward green. In this way, the habituative transmitter gates in a bottom-up pathway to the color ART circuit explain how typical afterimage arises from the gray stimulus after prolonged adaptation to its complementary color.

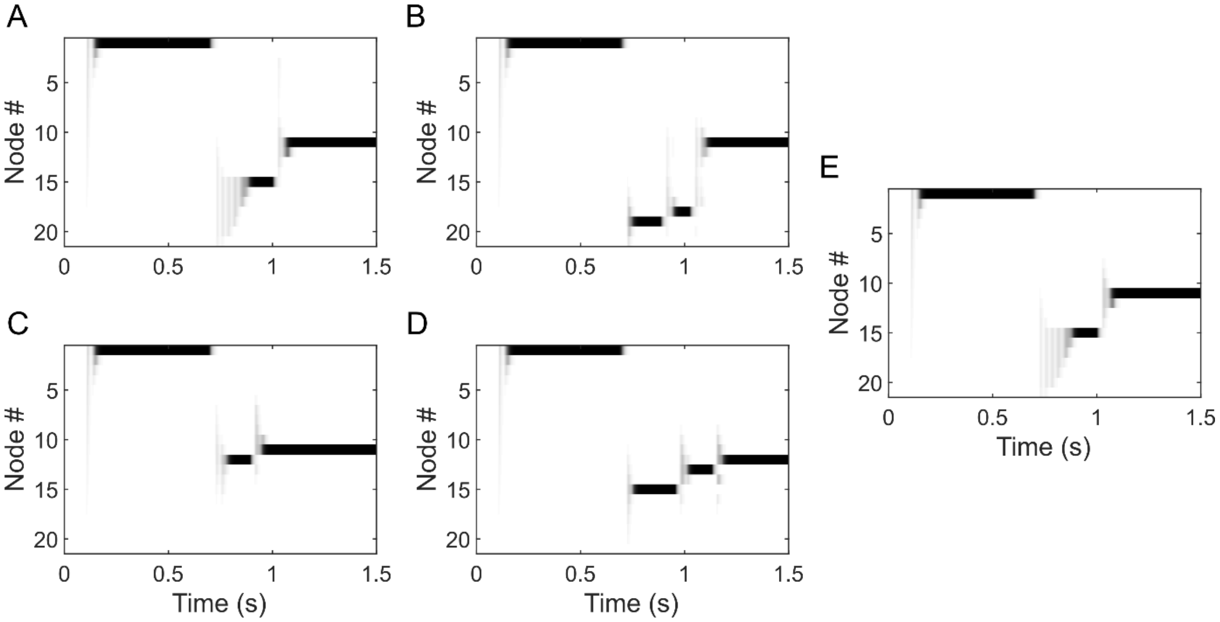


Figure 16. *Simulation of the Spanish castle illusion. The activity of the F2 layer is depicted in response to (A) the red-biased and (B) green-biased narrow feedback, as well as (C) the red-biased and (D) green-biased wide feedback, and (E) in a neutral condition without feedback.*

Figure 16 illustrated how feedback from the inter-ART associative map to the F_2 layer makes color sensation stronger (more vivid) in a feedback-enhanced afterimage. In all panels, the bottom-up input was always the same as depicted in Figure 15. Feedback was delivered to the F_2 layer from t = 600 ms until the end of simulation. Here, we varied the orientation of feedback distribution that could either induce bias to red or to green hues along with the type of feedback (narrow vs. wide). When the red-biased narrow feedback was applied, Node 15 was selected in the short temporal interval after recovery from the adaptation to pure red (Fig. 16A). The same node was also chosen as a winner in the control condition, that is, in a condition that generates a typical afterimage with no feedback present (Fig. 16E). This is consistent with the experimental report of Lupyan (2015b) who observed no statistically significant difference between a condition where bias was induced toward adapted hue and a control condition. On the other hand, the green-biased narrow feedback forced selection of Node 19 in the same temporal window (Fig. 16B). This node is tuned to the input of $I_L = 1$ and $I_M = 9$ corresponding to perception of more vivid greenish hue relative to Node 15, which is activated by typical afterimage. Node 19 is also closer to pure green encoded by Node 21. In this way, the green-biased narrow feedback increased vividness of the experience of a green color in the afterimage as observed in the Spanish castle illusion. As in previous simulations, we note that such a feedback-enhanced afterimage is fleeting, and it is removed from the F_2 layer by the activation of the orienting subsystem at around 1,100 ms. Therefore, our simulation also explains why we eventually arrive at the realization that we are watching an achromatic scene. Interestingly, informal observation suggests that this realization comes suddenly, akin to the effect of the reset signal on the F_2 layer.

Wide feedback with bias to red reduced color experience of the normal afterimage (Fig. 16C). Here, the F_2 layer chose Node 12 immediately after adaptation. This node is an immediate neighbor of Node 11 and corresponds to a weakest possible bias toward green. No such effect has been reported by Lupyan (2015b). However, it is possible that he selected only those objects that will produce a strong effect on the afterimage and avoided objects producing a weaker effect. We also observe that the green-biased wide feedback cannot generate the feedback-enhanced afterimage because its peak is located on Node 15 (Fig. 16D), the same node that is also selected in the control condition that generates a typical afterimage (Fig. 16E). Here, the only effect of feedback was to allow Node 15 to spend more time as a winner. We also note that this is the only simulation where the F_2 layer chose Node 12 instead of Node 11 as a final winner. The reason for this anomalous behavior is the fact that the total activity in the F_0 and F_1 was slightly lower than in the non-adapted state. Thus, the fuzzy intersection between the activity of F_0 and the top-down signals generated by Node 12 failed to cross the vigilance threshold to activate the orienting subsystem. However, it is reasonable to assume that transmitter gates would eventually recuperate from habituation and that they would reinstate veridical signal transmission. This would increase the discrepancy between the F_0 and F_1 activity enough to trigger reset of the F_2 layer.

7.4. Simulation of the Effect of Color Working Memory

Figure 17 illustrates how narrow (top row) and wide (bottom row) feedback from the inter-ART associative map bias the transfer of activity from the F_2 layer to the WM circuit. As in simulation 7.2., the bottom-up input was $I_L = 5.0$ and $I_M = 5.0$ from $t = 200$ ms until the end of simulation. Feedback was supplied in the same temporal interval. Suppose that the observer attends to the stimulus with a familiar shape first. In this case, object recognition in the shape ART circuit generated either narrow or wide feedback. The response of the F_2 layer to the feedback signals was the same as already observed in Figure 11B. In parallel, the WM circuit tracked the initial activity of the F_2 layer and selected the hue node at the similar spatial location. In the case of narrow feedback, the WM circuit selected Node 3, and in the case of wide feedback, it selected Node 7. Importantly, the WM circuit was sensitive to the F_2 input only at the early stage of the temporal evolution of its activity. Next, recurrent interactions within the WM circuit (i.e., lateral inhibition and self-excitation) started to dominate over the bottom-up input. Thus, the WM circuit remained insensitive to further changes that occurred later on in the F_2. Like the F_2 layer, the WM circuit is a hysteretic network but without external reset that can direct its activity toward the bottom-up input.

Figure 17 shows how a discrepancy in color representation is created between the F_2 and the WM circuit. Next, suppose that the observer moves its attention to another gray stimulus without recognizable shape and judges its color. In this case, there would be no feedback signals from the inter-ART associative map. Thus, the F_2 layer would immediately select Node 11 while the WM circuit would maintain its previously established hue representation. When asked to make a color judgment, the observer will judge the color of the currently attended stimulus in relation to the color representation held in WM. This would produce a bias because a previously attended object would be judged as more reddish than it actually is. For example, a heart is judged to be more reddish than a circle of the same hue as observed in the study of Delk and Fillenbaum (1965) and in experiment 3 of Valenti and Firestone (2019). The same processes described above may also help to explain the effect of categorization of colored letter and numerals on hue perception (Goldstone, 1995) and the effect of positive and negative emotions on brightness perception (Banerjee et al., 2012; Song et al., 2012). More generally, such WM bias may contaminate the results of all experimental tasks involving comparison between two or more stimuli.

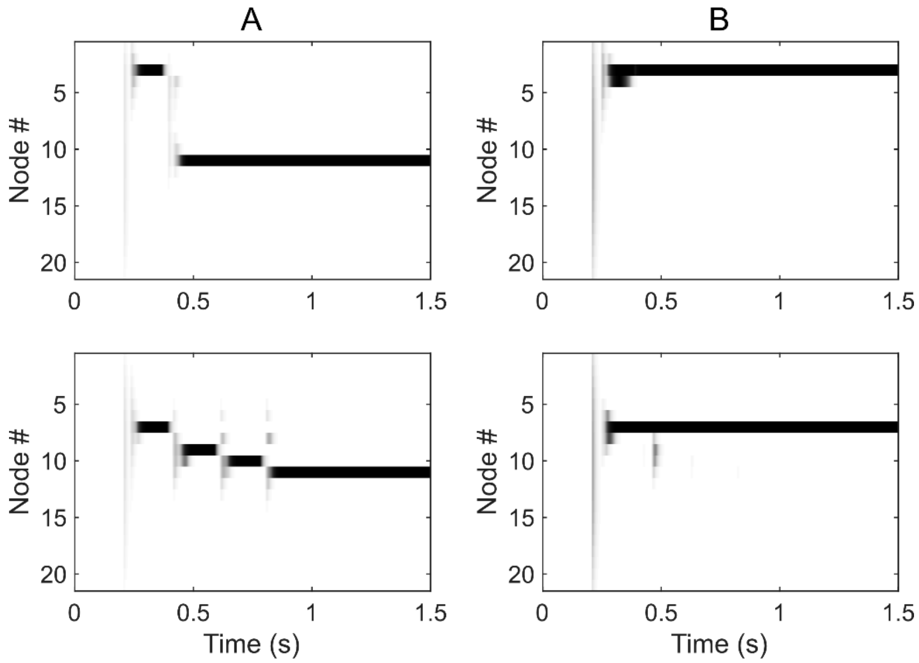


Figure 17. *Transfer of activity from (A) the F2 layer to (B) the color working memory circuit when (top row) the red-biased narrow or (bottom row) red-biased wide feedback was applied to the F2 layer.*

7.5. Simulation of Neural Evidence for Cognitive Penetrability of Color Vision

We also examine the fMRI evidence of CPV obtained when participants view the achromatic images of objects with a diagnostic or typical color (Bannert & Bartels, 2013; Vandenbroucke et al., 2016). An important feature of the BOLD signal is that it represents the spatiotemporal average of activity of a large number of neurons taken over an extended period of time (on the order of several minutes). The dynamics of blood flow are slower relative to neural events that trigger its local increase or decrease. For this reason, the BOLD signal always represents a delayed and blurred picture of neural activity. In this regard, our explanation of the fMRI studies demonstrating CPV follow the same logic as an explanation of behavioral findings. Figure 18 illustrates this point. We computed metabolic response of the F_2 nodes as an integral of their spatially blurred activity. The integral was taken over the temporal interval of one second. Here, we considered four conditions: red and gray color without feedback and gray color with narrow or wide feedback. Bottom-up input and narrow or wide feedback were delivered from t = 200 ms to the end of simulation at t = 1,000 ms. The metabolic response to a gray color was clearly biased toward red in the presence of either narrow or wide feedback.

Thus, it would be possible to decode the brain signal associated with an expected color despite the fact that the observer is actually perceiving achromatic color.

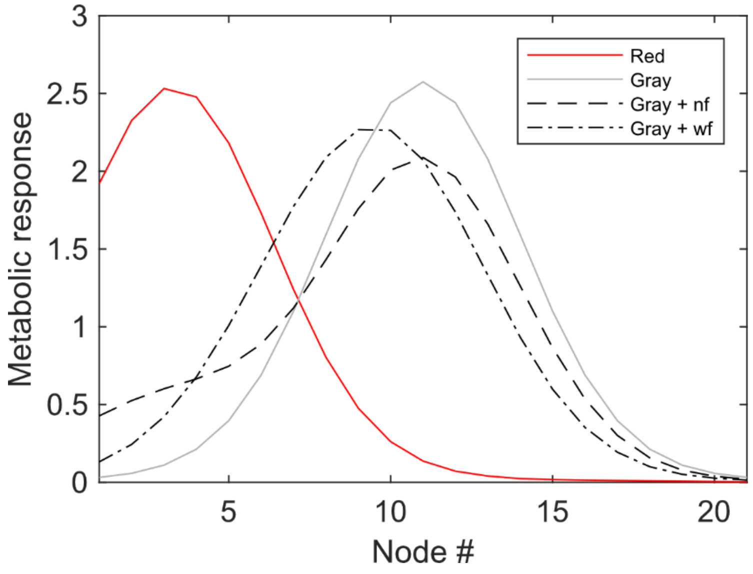


Figure 18. *Simulated metabolic response of the F2 layer when the input was a red (IL = 9.0, IM = 1.0) or gray color (IL = 5.0, IM = 5.0) without feedback and when the gray color was presented together with the red-biased narrow feedback (NF) or the red-biased wide feedback (WF).*

In addition, there is evidence that the expected color signal can also be decoded in V1 (Bannert & Bartels, 2013). Given that color tuning is not present in V1, it is not likely that such activity represents color perception (Conway et al., 2007; Conway & Tsao, 2009). Rather, it may reflect the activation of the global feature-based attention that gives a competitive advantage to the expected color. In other words, the inter-ART associative map may not only send feedback to the F_2 layer to create an expectation of what is likely to occur next, but also create a competitive bias in the early stages of processing where it may facilitate the registration and transmission of an expected feature value. Still, this does not mean that V1 is penetrable. We argue that feature-based attention as well as object- and space-based attention operate via a separate network. This will be further elaborated in Section 8.1. *Adaptive resonance theory and attention*.

8. SPATIAL ATTENTION

8.1. Adaptive Resonance Theory and Attention

Lupyan (2017a) has argued that directing attention to a part of a visual scene constitutes evidence of CPV because moving attention to a location in space effectively alters appearance at that location. In addition, numerous studies of single-unit recordings in the monkey brain revealed that attention modulates response properties of neurons in the visual cortex (Reynolds & Chelazzi, 2004; Treue, 2001). In this section, we address the question of how attention interacts with the color ART circuit without disrupting its ability to encode hue of the bottom-up input. Following Beck and Schneider (2017), and motivated by the anatomical findings about segregated feedforward and feedback pathways in the visual cortex (Markov et al., 2013, 2014; Markov & Kennedy, 2014), we suggest that early stages of visual processing consist of two parallel and segregated networks, which are labeled as a saliency network and a veridical network (Fig. 19).The saliency network is dedicated to the computation of saliency based on the bottom-up or top-down cues. It culminates in the selection of all locations occupied by a single object that is in the current focus of attention. On the other hand, the veridical network is dedicated to faithful transmission of feature values registered in the retina and passed through the LGN and V1 to higher-level centers such as inferotemporal cortex. From this view, all available evidence of the attentional modulation of neurons in the LGN, V1, and V2 actually reveals the operation of the saliency network. However, in parallel, a separate veridical network exists that encodes feature values (i.e., the amount of redness or greenness) at all locations uncontaminated by top-down influences. To sample the specific feature value that occurs on the attended object, the saliency network interacts with the veridical network in V4. One way in which to achieve this is via dynamic routing by multiplicative gating of the signal flow from V2 to V4 (Heinke & Humphreys, 2003; Olshausen et al., 1993, 1995). Another possibility is to induce synchronization of oscillatory activity specifically for the object of attention (Baldauf & Desimone, 2014). The source signal for this routing may arise from the posterior parietal cortex (PPC) as shown by Van Dromme et al., (2016).

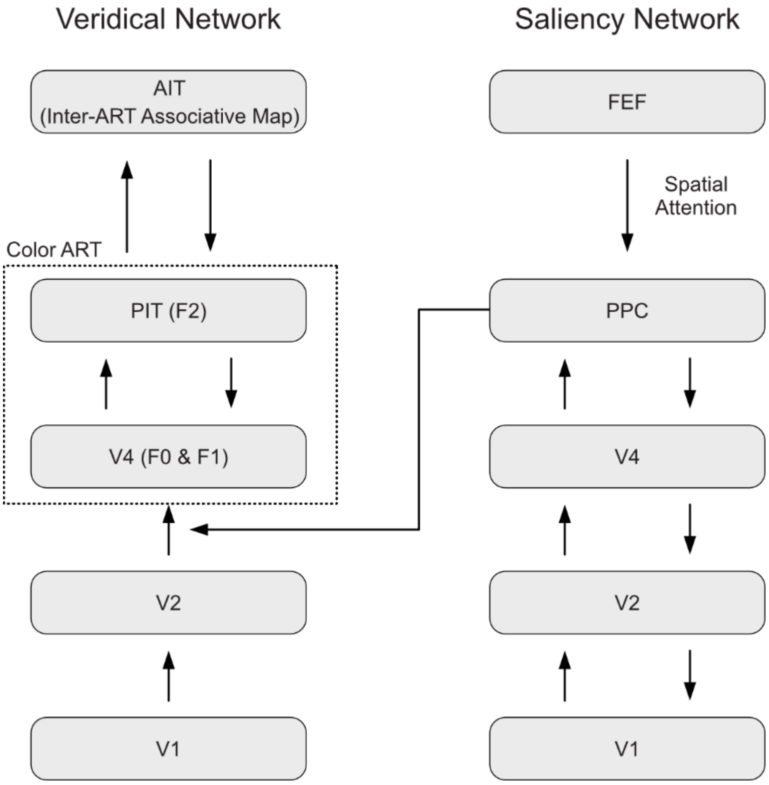


Figure 19. *Segregation between parallel veridical and saliency networks in the early visual cortex explains how attention might select an object without disrupting the color-opponent signals used to compute hue in the PIT.*

To select the attended feature value without destroying the hue information it carries, the PPC should operate as a feature-based WTA network. It should contain a binary spatial representation akin to a Boolean map proposed by Huang and Pashler (2007). In this map, all locations belonging to the attended object are marked by an elevated firing rate, and all other locations are suppressed to zero. Marić and Domijan (2018) showed how a WTA network may implement the Boolean map. A similar idea is expressed by the concept of the attentional shroud, which covers the attended object and suppresses its background (Fazl et al., 2009; Foley et al., 2012). Next, the output of the PPC interacts multiplicatively with the feedforward signal flow from V1 or V2 to V4. Multiplication assures that the color signal in V4 arriving from the attended object remains the same as if the object is presented alone in the scene. On the other hand, unattended locations are filtered out because they are multiplied by zero. These processes are illustrated in Figure 20. Another possibility is that attention segregates distinct perceptual groups into separate segmentation layers in V2 (Francis et al., 2017). In this model, each segmentation layer operates as a distinct Boolean map. In either way, color signals can be dynamically routed from retinotopic maps in V1 to translation-invariant color representation in V4 (Heinke & Humphreys, 2003; Olshausen et al., 1993, 1995).

Such a proposal is consistent with a biased competition model and empirical studies suggesting that attention in V4 operates as a filter, which leaves the representation of the attended object intact and suppresses representation of distractors (Desimone & Duncan, 1995; Reynolds et al., 1999). A similar idea has been used in a model of interaction between saliency computation and object recognition (Walther & Koch, 2006). The proposed model uses a spatial representation similar to the Boolean map to encode all locations occupied by the selected object. This information is then projected back to multiplicatively gate the signal flow in the object recognition network.

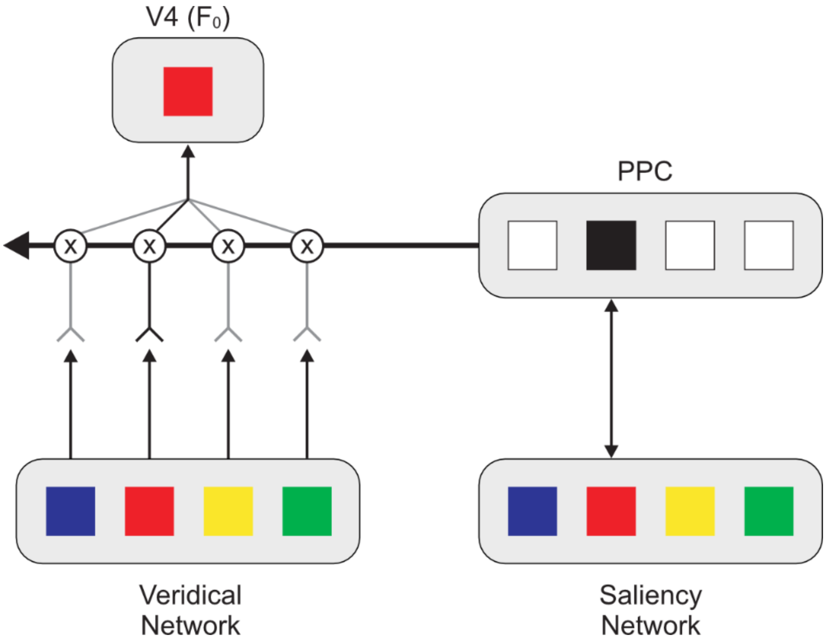


Figure 20. *Detailed view of the interaction between the PPC and V4 that generates translation-invariant input to V4. In this example, input consists of four colored squares, and spatial attention is directed to the second square from the left. The selected square is depicted in black, while suppressed squares are marked in white in the PPC, and X denotes multiplication occurring on dendrites of the V4 nodes.*

After the cone-opponent signal of the attended object is registered in V4, the color ART circuit may select a color category in the PIT that best matches with the bottom-up activity. As part of this process, feedback projections from the PIT to V4 read out the top-down, learned expectations in V4. Furthermore, the PIT may receive its own top-down signals from the AIT where the inter-ART associative map integrates information about the object’s color and shape.

Therefore, the AIT → PIT → V4 is a feedback pathway responsible for communicating predictions to V4. Importantly, this feedback is segregated from the feedback pathway between the PPC and V1, which is involved in spatial selection. Finally, it should be noted that the computation in the saliency network does not need to be restricted to spatial attention. Instead, it may be guided by a chosen non-spatial feature value such as a red color. In addition, spatial and feature-based attention may interact together to specify the location of the target stimulus (Leonard et al., 2015; van Es et al., 2018).

The proposed scheme for the interaction between spatial attention and the color ART circuit clarifies why attention and expectations should be treated as separate sources of feedback signals (Summerfield & de Lange, 2014; Summerfield & Egner, 2009, 2014). The present analysis suggests that attention and expectations solve different computational problems. First, spatial attention is needed to solve the superposition catastrophe or feature-binding problem in the multi-object scene (von der Malsburg, 1999). In this regard, spatial attention is required to highlight features belonging to the attended object and to bind them together into a unified object representation. At the same time, features of the unattended objects are filtered out by suppressing their neural representation. In this regard, spatial attention must operate before the color ART circuit begins to encode the color-opponent inputs into a perceived hue. Second, expectations are needed to solve the stability–plasticity dilemma and to prevent catastrophic forgetting, as discussed above (Grossberg, 1980).

8.2. ARTSCAN

The previous section presented a simplified view on the possible interactions between spatial attention and the ART circuit. An alternative and more detailed explanation of the way in which spatial attention and eye movements interact with translation-invariant object recognition during active vision is provided in an ARTSCAN model developed by Grossberg and colleagues (Cao et al., 2011; Chang et al., 2014; Fazl et al., 2009; Foley et al., 2012). It consists of two parallel processing streams – the *What Stream* and the *Where Stream* – mimicking the division of labor between ventral and dorsal pathways in the visual cortex (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). On the one hand, the What Stream is associated with object learning, recognition, and prediction. It progressively builds a more abstract object representation from view-specific nodes in the PIT to view-invariant nodes in the AIT. On the other hand, the Where Stream is involved in object localization, spatial- and object-based attention, and predictive remapping during eye movements. The model explains how view-specific and view-invariant representation of object categories develops during active exploration of the visual scene.

Based on the ARTSCAN model, Grossberg (2017b) has suggested that visual consciousness can be divided into two distinct experiences supported by two different types of resonance (Fig. 21A). The first is an experience of recognizing an object or a scene that is driven by feature-category resonance, While the second is an experience of seeing a visual object or a scene driven by surface-shroud resonance. Feature-category resonance was already discussed and involves a ventral visual stream with its recognition codes in the AIT cortex. In contrast, surface-shroud resonance develops in the recurrent pathway between V4 and the PPC when we selectively attend to an object. The input to this process is the surface representation in V4. It receives feedback from the PPC, which selectively amplifies locations and features belonging to the attended object and suppressed locations and features of unattended objects. In this way, surface-shroud resonance modulates the input to feature-category resonance and prevents simultaneous encoding of features of multiple objects. These two types of resonance are coordinated to bring the unified conscious experience of what we are looking at and where it is.

Grossberg (2017b) has suggested that the distinction between surface-shroud and feature-object resonance reflects the distinction between seeing and knowing. In contrast, our proposal for color coding in the PIT rests on the idea that surface-shroud resonance is a prerequisite, but it is not sufficient to generate color qualia (Fig. 21B). Surface-shroud resonance enables the selection of a set of locations from which feature values will be extracted. When certain locations are selected, feature values such as color-opponent responses at those locations become available to enter into the color and shape ART circuits. In other words, we suggest that independent feature-category resonances develop for colors and shapes (and possibly other perceptual attributes) in the PIT. Such parallel feature-category resonances in the color and shape ART circuits correspond to seeing an object’s color and shape, respectively. At the next processing stage in the hierarchy (AIT), color and shape are bound together into a unified object representation that gives rise to conscious knowledge about the object.

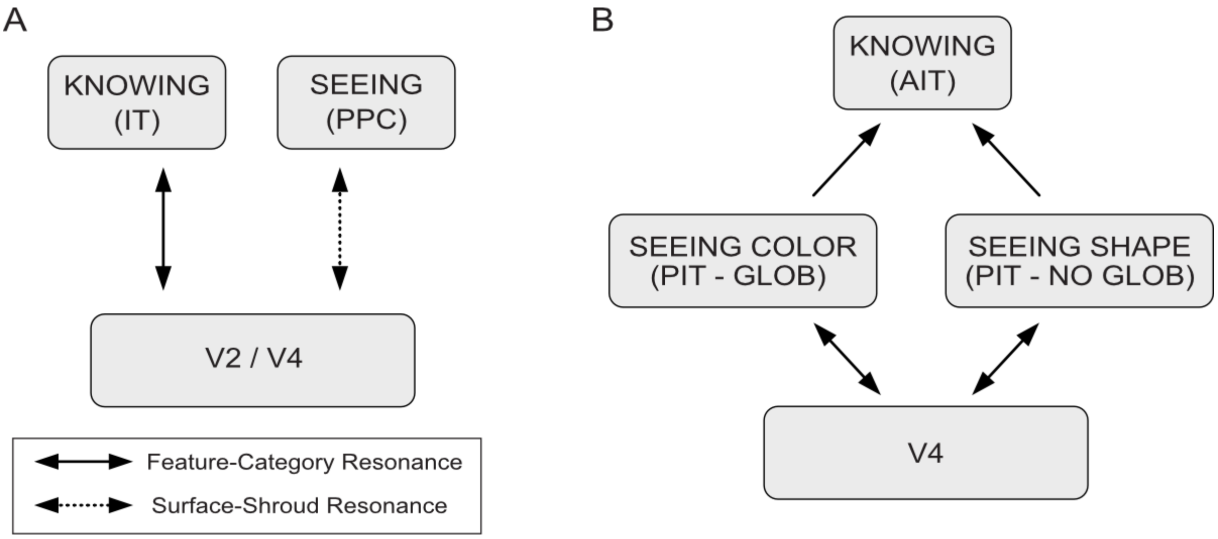

A
KNOWING (IT)
SEEING (PPC)
V2 / V4
Feature-Category Resonance
Surface-Shroud Resonance
B
KNOWING (AIT)
SEEING COLOR (PIT - GLOB)
SEEING SHAPE (PIT - NO GLOB)
V4

Figure 21. *(A) Distinction between seeing and knowing based on their neural circuitry, according to Grossberg (2017b). (B) The modification proposed here is based on recent findings regarding the neural underpinnings of color perception (Conway et al., 2007; Conway, 2009; Conway & Tsao, 2009).*

9. CONCLUSION

In this work, we showed through computer simulations that the ART circuit is a specific computational implementation of the predictive coding system that is not penetrable by top-down influences. A unique feature of the ART circuit is that it is an attractor neural network supplemented by two control mechanisms: gain control and an orienting subsystem. Their role is to solve the stability–plasticity dilemma, that is, to simultaneously prevent the erosion of previously learned recognition codes and enable sufficient plasticity to learn new codes in the future. Balance between stability and plasticity is achieved by constraining the impact of top-down predictions on the ongoing network processing. As a side effect of the operation of gain control and the orienting subsystem, conscious visual perception is almost never cognitively penetrable. The ART circuit is vulnerable to top-down influences during the period of searching through the state space to find the best match between bottom-up input and the choice of hue category. Traces of this memory search are observed in behavioral and functional neuroimaging studies. Crucially, the ART search cycle is a transient process. After the network reaches the attractor, it is fully resistant to any further top-down influence. In other words, the ART circuit enters into a resonant state that corresponds with conscious visual perception.

ACKNOWLEDGMENT

This work was supported by the University of Rijeka under Grant uniri-drustv-18-177.

APPENDIX. Model Equations and Parameters

The model is defined by a set of ordinary differential equations, which are numerically solved using Euler's forward method. The time scale of the nodes' activities is taken to be a unit of time, with the convention that one unit of time is equal to 10 ms in real time. The step size of 0.01 used in numerical integration thus corresponds to 0.1 ms in real time. The Matlab code to reproduce all the results reported in this article have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/zphyq/.

A.1. Color Pre-processing

The activity of cone c_i with spectral sensitivity $i \in \{L, M\}$ is governed by the following equation:

$$\tau_c \frac{d}{dt} c_i = -c_i + I_i \,. \tag{1}$$

Term τ_c in Eq. (1) denotes the cone's time constant, and I_i is an input amplitude (light intensity) selected from the following set:

$$I_i \in \{0.0, 0.5, 1.0, 1.5, \dots, 10.0\} \tag{2}$$

with the constraint that $I_M = 10.0 - I_L$. The activities of the L- and M-cones thus complement each other. Each cone projects its output to the F_0 node with the same spectral selectivity. The activity of F_0 node $x_i^{(0)}$ is governed by

$$\tau_x \frac{d}{dt} x_i^{(0)} = -x_i^{(0)} + \frac{z_i c_i}{\alpha_0 + y^{(0)}} \tag{3}$$

where τ_x is a time constant of the excitatory node, and α_r controls the slope of its activation as a function of input amplitude. Cones also project their output to a common inhibitory interneuron $y^{(0)}$. Its activity is defined by

$$\tau_y \frac{d}{dt} y^{(0)} = -y^{(0)} + \sum_i z_i c_i \tag{4}$$

where τ_y is a time constant of the inhibitory node. An inhibitory interneuron normalizes the activity of F_0 nodes to the interval [0, 1]. Furthermore, excitatory and inhibitory projections are both modulated by habituative transmitter gate z_i. Temporal evolution of the transmitter gate is governed by the following equation:

$$\tau_z \frac{d}{dt} z_i = \beta\left(\gamma - z_i\right) - \delta z_i c_i \tag{5}$$

where τ_z is a time constant of transmitter habituation, and β controls the rate of transmitter production until it reaches upper limit γ. Parameter δ controls the rate of transmitter depletion triggered by the presynaptic activity arriving from cone c_i. Moreover, the parameter values of the color pre-processing circuit were set as τ_c = 50 ms, $\tau_x = \tau_y$ = 10 ms, τ_z = 1,500 ms, α_0 = 0.001, β = 0.1, and γ = 1. Parameter δ was set to 10 in the simulation presented in Section 7.3. involving afterimages, and δ = 0 in all other simulations.

A.2. Color ART Circuit

In the fuzzy ART, the F_1 layer computes the fuzzy intersection between the vector of bottom-up input and the vector of top-down weights that are read out by the winning node in F_2. Fuzzy intersection is defined as a component-wise minimum between two vectors. The activity of F_1 node $x_i^{(1)}$ with spectral sensitivity i is thus defined by

$$\tau_x \frac{d}{dt} x_i^{(1)} = -x_i^{(1)} + g_1 x_i^{(0)} + \left(1 - g_1\right) \min\left(x_i^{(0)}, X_i^{(2)}\right) \tag{6}$$

where $X_i^{(2)}$ denotes the total postsynaptic current arriving from the top-down pathway $F_2 \rightarrow F_1$ given by

$$X_i^{(2)} = \sum_j \min\left(w_{ji}, x_j^{(2)}\right). \tag{7}$$

Eq. (7) describes the read-out of top-down adaptive weights generated by fuzzy intersection between the weight vector and the vector of F$_2$ activities.

The problem with fuzzy intersection in Eq. (6) is that the F$_1$ layer will remain inactive if no suprathreshold top-down signals are present. To address this issue, we designed a gain control node g_1 to allow the F$_1$ node to pass bottom-up activation to the F$_2$ layer in the absence of top-down signals. This is achieved by multiplicative gating of the activity of the F$_1$ layer by g_1. The g_1 node receives excitation from the F$_0$ layer and inhibition from the F$_2$ layer. The action of g_1 on the F$_1$ layer mimics the effect of acetylcholine on the cells in the visual cortex (Harris & Thiele, 2011). On the one hand, elevated levels of acetylcholine strengthen feedforward signal transmission and reduce the impact of feedback signals. On the other hand, a lower level of acetylcholine has an opposite effect: it strengthens the impact of feedback signals on the cell's activity. In the same manner, when g_1 is active, F$_1$ is able to transmit bottom-up signals from F$_0$ to F$_2$ without top-down signals. When g_1 is inactive, top-down signals are read out in F$_1$, and F$_1$ computes a fuzzy intersection.

The activity of gain control node g_1 is defined by

$$\tau_g \frac{d}{dt} g_1 = -g_1 + \left[H\left(\sum_i x_i^{(0)} - \theta_0 \right) - H\left(\sum_j x_j^{(2)} - \theta_2 \right) \right]^+ \tag{8}$$

where τ_g is a time constant, $[u]^+$ denotes the threshold-linear or half-wave rectification function

$$[u]^+ = \max(u, 0) \tag{9}$$

and $H(u)$ denotes a Heaviside step function

$$H(u) = \begin{cases} 0 & if \quad u \le 0 \\ 1 & if \quad u > 0 \end{cases}. \tag{10}$$

Thresholds θ_0 and θ_2 in Eq. (8) prevent the hair-trigger activation of g_1 by low activity levels in the F$_0$ and F$_2$ layers, respectively.

The activity of F$_2$ node $x_j^{(2)}$ at spatial location $j \in \{1, \dots, N\}$ is defined by

$$\tau_x \frac{d}{dt} x_j^{(2)} = -x_j^{(2)} + S\left(\sum_i \min\left(w_{ij}, x_i^{(1)}\right) + g_2\left(\varepsilon_2 x_j^{(2)} + a_j + b_j\right) - y^{(2)} - \chi y_j^{(3)} \right) \tag{11}$$

where *S*(*u*) is a piecewise-linear function

$$S(u) = \begin{cases} 0 & if \quad u \le 0 \\ u & if \quad 0 < u < \lambda \\ \lambda & if \quad u \ge \lambda \end{cases} \tag{12}$$

with upper saturation point λ. The bottom-up input from the F_1 layer is passed through a filter of synaptic weights w_{ij}. The bottom-up pathway also computes fuzzy intersection in order to generate a postsynaptic current on the F_2 node. The activity of F_2 gain control node g_2 is defined by

$$\tau_g \frac{d}{dt} g_2 = -g_2 + H\left(\sum_i x_i^{(0)} - \theta_0 \right) \tag{13}$$

where θ_0 is a threshold. The g_2 node multiplicatively gates all sources of feedback signals to the F_2 node, including its self-excitation whose strength is controlled by ε_2.

Term a_j denotes the spatial gradient of intrinsic activity that was used in the simulation of unsupervised learning (simulation presented in Section 7.1.). Intrinsic activity is defined by

$$a_j = 1 - 0.05(j-1). \tag{14}$$

In all other simulations, $a_j = 0$ for all *j*.

Feedback signals from the inter-ART associative map to the F_2 node was modeled as

$$b_j = \phi\left[1 - \eta^{-1}|\mu - j|\right]^+ \tag{15}$$

where ϕ is a feedback gain factor, μ determines the location of the peak of the feedback distribution, and η determines its spread. Narrow feedback with a bias toward red (green) hues was defined by $\mu = 3$ ($\mu = 19$) and $\eta = 5$. Wide feedback with a bias toward red (green) hues was defined by $\mu = 7$ ($\mu = 15$) and $\eta = 10$. The feedback gain factor was set at $\phi = 0.3$ in both narrow and wide distribution. In the simulation presented in Figure 11, the feedback gain factor was drawn from the set $\phi \in \{0.1, 0.3, 0.5\}$. Feedback signals were set to $b_j = 0$ for all j during learning (simulations presented in Section 7.1.).

In Eq. (11), the F$_2$ node receives two sources of inhibition. One source is an inhibitory interneuron $y^{(2)}$, mediating lateral inhibition that leads to a WTA choice among the F$_2$ nodes. The activity of $y^{(2)}$ is governed by the following equation:

$$\tau_y \frac{d}{dt} y^{(2)} = -y^{(2)} + \sum_j x_j^{(2)} . \tag{16}$$

Another source of inhibition is a reset signal mediated by a separate inhibitory interneuron $y_j^{(3)}$. Its impact on the F$_2$ node is controlled by the parameter χ. The activity of the inhibitory interneuron carrying the reset signal is governed by the following equation:

$$\tau_y \frac{d}{dt} y_j^{(3)} = -y_j^{(3)} + x_j^{(3)} . \tag{17}$$

Interneuron $y_j^{(3)}$ receives excitation from the F$_3$ node (Fig. 22). Its role is to enable a search for the best-matching F$_2$ node. The layer of F$_3$ nodes continues to inhibit all F$_2$ nodes that are already visited during a single stimulus presentation. In this way, the F$_3$ layer operates as an interface between the attentional and the orienting subsystem (Domijan & Šetić, 2016). It allows a non-specific reset signal generated in the orienting subsystem to be delivered specifically to the currently winning F$_2$ node. The activity of F$_3$ node $x_j^{(3)}$ is defined by

$$\tau_x \frac{d}{dt} x_j^{(3)} = -x_j^{(3)} + S\left(\left[x_j^{(2)} + r - \theta_3\right]^+ + g_2 \varepsilon_3 x_j^{(3)}\right) \tag{18}$$

where $S(\cdot)$ is a node's output function, $[\cdot]^+$ is a half-wave rectification describing computation at the dendrite of the F$_3$ node, and θ_3 is a dendritic threshold. The dendrite of the F$_3$ node detects coincident activation of the F$_2$ node at the same network location j and of the reset node r defined by Eq. (19). Furthermore, the F$_3$ node is endowed with a self-recurrent collateral that enables it to sustain suprathreshold activity after the corresponding F$_2$ node is inhibited. The strength of self-excitation is controlled by the parameter ε_3. In addition, self-excitation is multiplicatively gated by the gain control *g*$_2$ in order to abolish sustained activity of the F$_3$ node when the stimulus disappears. Such gating enables the F$_3$ layer to start a new memory search when new input is presented.

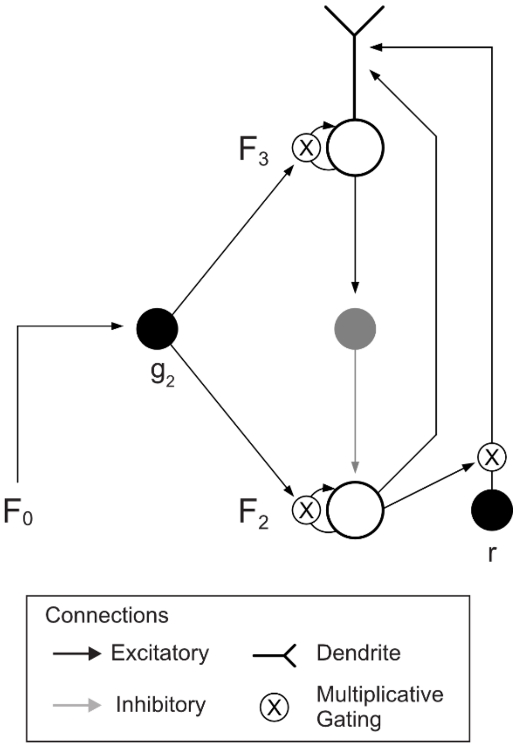


Figure 22. *A node in the F3 layer with connections to other model components.*

The degree of match between the total activity in the F$_0$ and F$_1$ layers is monitored by the reset node *r*, whose activity is governed by the following equation:

$$\tau_r \frac{d}{dt} r = -r + H\left(\sum_j x_j^{(2)} - \theta_r\right) H\left(\frac{\sum_i x_i^{(1)}}{\alpha_r + \sum_i x_i^{(0)}} < \rho\right) \tag{19}$$

where τ_r is a time constant of the *r* node that is chosen to be slower than the time constants of the excitatory nodes in the F$_0$ and F$_1$ layers ($\tau_r \gg \tau_x$) to make the *r* node less sensitive to transient fluctuations in the F$_0$ and F$_1$ activity. The second term on the r.h.s. of Eq. (19) describes a gating of the *r* node by the activity of the F$_2$ layer. The term θ_r is a threshold that prevents the activation of the *r* node when the F$_2$ layer activity is weak. The last term on the r.h.s. of Eq. (19) describes a test of how much smaller the total activity in the F$_1$ layer is than the total activity in the F$_0$ layer, where α_r is a choice parameter and ρ is vigilance. Gating by the F$_2$

activity was introduced to prevent premature activation of the orienting subsystem. On the first wave of the signal flow through the attentional subsystem, the F_0 layer is activated before F_1. This creates a short temporal window when a mismatch occurs between the activity of F_0 and F_1 that is not a consequence of the mismatch between sensory and top-down signals in F_1 because F_1 is not yet activated by the F_0. Gating by the F_2 activity consequently assures that the *r* node will become active only when two events occur simultaneously: there is a mismatch between the size of total activity in the F_0 and F_1 layers, and there is a suprathreshold activity in the F_2 layer.

The parameter values of the color ART circuit were set as $\tau_x = \tau_y = \tau_g$ =10 ms, τ_r = 100 ms, N = 21, λ = 2, $\varepsilon_2 = \varepsilon_3$ = 2, χ = 2, θ_0 = 0.1, θ_2 = 1.5, θ_3 = 2.5, θ_r = 1.5, α_r = 0.001, and ρ = 0.96. In the simulation presented in Figure 13, the time constant of the F_2 node was drawn from the set $\tau_x \in \{25, 50, 100\}$. In the simulation presented in Figure 14, vigilance was drawn from the set $\rho \in \{0.5, 0.7, 0.9\}$.

A.3. Learning in the Color ART Circuit

Synaptic weight w_{ij} in the feedforward pathway F_1 → F_2 is updated according to the following equation:

$$\tau_w \frac{d}{dt} w_{ij} = H\left(x_j^{(2)} - \theta_w - q_j\right)\left[\min\left(w_{ij}, x_i^{(1)}\right) - w_{ij}\right] \qquad (20)$$

where τ_w is a time constant of synaptic change that may be defined as an inverse of a learning rate parameter typically used in neural models of learning. Parameter θ_w is a static threshold, and q_j is a dynamic threshold defined in Eq. (21). As before, function minimum denotes a fuzzy intersection between the top-down weight vector and the F_1 activity vector.

Learning occurs in a fast, one-shot manner; that is, one stimulus presentation is sufficient to encode it into a weight vector of the winning F_2 node. The second term on the r.h.s. of Eq. (20) describes a synaptic gate that is open only when F_2 crosses the static and dynamic threshold. The static threshold is set to a high value, close to the maximal activity level of the F_2 node. In this way, the static threshold assures that weight adaptation occurs only at the synapses of the winning F_2 node. When the synaptic gate is open, the weights of the winning F_2 node approach fuzzy intersection between the F_1 activity vector and weight vector.

The temporal evolution of dynamic threshold q_j in Eq. (20) is described by

$$\tau_q \frac{d}{dt} q_j = H\left(x_j^{(2)} - \theta_q\right)\left(1 - q_j\right) \tag{21}$$

where τ_q is a time constant of the dynamic threshold with the constraint that $\tau_q \gg \tau_w$ to assure that the dynamic threshold elevates only after synaptic weights are adjusted. In this way, the dynamic threshold slowly elevates until it closes the synaptic gate in Eq. (20) and prevents further change in synaptic weights despite the fact that a stimulus may still be present. In this way, the dynamic threshold prevents later events from distorting the learned weights of the winning F$_2$ node. That is, after a reset of the current F$_2$ winner, a short period of time exists when its activity amplitude decays to zero. During this decay period, there is danger of distorting the weight vector of the winning F$_2$ node. The dynamic threshold prevents this from happening.

In the fuzzy ART, bottom-up and top-down adaptive weights are symmetric. Synaptic weight w_{ji} in the top-down pathway F$_2$ → F$_1$ is thereby updated according to equation analogous to Eq. (20):

$$\tau_w \frac{d}{dt} w_{ji} = H\left(x_j^{(2)} - \theta_w - q_j\right)\left[\min\left(w_{ji}, x_i^{(1)}\right) - w_{ji}\right]. \tag{22}$$

Synaptic weights in both pathways were initially set to their maximal values

$$w_{ij}(0) = w_{ji}(0) = 1, \quad \forall i, \forall j \tag{23}$$

and dynamic thresholds were set to

$$p_j(0) = 0, \quad \forall j. \tag{24}$$

Parameter values for the learning equations were set as τ_w = 20 ms, τ_q = 200 ms, and $\theta_w = \theta_q$ = 1.5.

A.4. Working Memory Circuit

The activity of the F$_2$ layer is passed forward to the working memory circuit labeled as an F$_4$ layer. It has the same anatomical arrangement as in the F$_2$ layer; that is, the F$_4$ layer consists of a set of excitatory nodes that are reciprocally connected to an inhibitory interneuron. The activity of F4 node $x_j^{(4)}$ is governed by the following equation:

$$\tau_x \frac{d}{dt} x_j^{(4)} = -x_j^{(4)} + S\left(\sum_k D(j,k) x_k^{(2)} + \varepsilon_4 x_j^{(4)} - y^{(4)} \right). \tag{25}$$

In Eq. (25), parameter *ε4* controls the strength of self-excitation, and distance-dependent function *D*(*j*, *k*) is defined by

$$D(j,k) = \left[1 - \nu^{-1} |k - j|\right]^+ \tag{26}$$

where *ν* controls its spatial spread. The activity of inhibitory interneuron $y^{(4)}$ is governed by the following equation:

$$\tau_y \frac{d}{dt} y^{(4)} = -y^{(4)} + \sum_j x_j^{(4)} . \tag{27}$$

The parameter values of the working memory circuit were set as $\tau_x = \tau_y = 10$ ms, $\varepsilon_4 = 2$, and $\nu = 20$.

A.5. Computation of the Metabolic Response

Metabolic demand m_j generated by the neural activity of the F2 node at location *j* is computed as

$$m_j = \int_{t_0}^{t_1} e_j(t)\, dt \tag{28}$$

where the integral is taken over the interval from $t_0 = 0$ to $t_1 = 1{,}000$ ms, and

$$e_j(t) = \sum_k G(j,k) x_k^{(2)}(t) \tag{29}$$

is a convolution of the F_2 layer activity with one-dimensional Gaussian function $G(j, k)$ defined by

$$G\left(j,k\right)=\frac{1}{2\pi\sigma^2}\exp\left[-0.5\frac{\left(k-j\right)^2}{\sigma^2}\right] \tag{30}$$

where $\sigma = 3$ controls the spatial spread of the Gaussian function.

REFERENCES

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences,* 8(10), 457–464. https://doi.org/fh8cch

Allred, S. R., & Olkkonen, M. (2015). The effect of memory and context changes on color matches to real objects. *Attention, Perception & Psychophysics*, *77*(5), 1608–1624. https://doi.org/10.3758/s13414-014-0810-4

Ashby, F. G., & Rosedahl, L. (2017). A neural interpretation of exemplar theory. *Psychological Review*, *124,* 472–482. https://doi.org/10.1037/rev0000064

Asplund, C. L., Fougnie, D. L., Zughni, S., Martin, J., & Marois, R. (2014). The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychological Science*, *25*(3), 824–831. https://doi.org/10.1177/0956797613513810

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J. R., & Thierry, G. (2010). Perceptual shift in bilingualism: brain potentials reveal plasticity in preattentive colour perception. *Cognition, 116*(3), 437–443. https://doi.org/10.1016/j.cognition.2010.05.016

Bacigalupo, F., & Luck, S. J. (2015). The allocation of attention and working memory in visual crowding. *Journal of Cognitive Neuroscience, 27*(6), 1180–1193. https://doi.org/10.1162/jocn_a_00771

Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision, 14*(4), 7. https://doi.org/10.1167/14.4.7

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science, 344*(6182), 424–427. https://doi.org/10.1126/science.1247003

Banerjee, P., Chatterjee, P., & Sinha, J. (2012). Is it light or dark? Recalling moral behavior changes perception of brightness. *Psychological Science, 23,* 407–409. https://doi.org/10.1177/0956797611432497

Bannert, M. M., & Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology, 23*(22), 2268–2272. https://doi.org/10.1016/j.cub.2013.09.016

Baraldi, A., & Alpaydin, E. (2002a). Constructive feedforward ART clustering networks. I. *IEEE Transactions on Neural Networks, 13*(3), 645–661. https://doi.org/10.1109/TNN.2002.1000130

Baraldi, A., & Alpaydin, E. (2002b). Constructive feedforward ART clustering networks. II. *IEEE Transactions on Neural Networks, 13*(3), 662–677. https://doi.org/10.1109/TNN.2002.1000131

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.–M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., & Fries, P. (2015). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. *Neuron, 85*(2), 390–401. https://doi.org/10.1016/j.neuron.2014.12.018

Beck, J., & Schneider, K. A. (2017). Attention and mental primer. *Mind & Language, 32*(4), 463–494. https://doi.org/10.1111/mila.12148

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567–575. https://doi.org/10.1016/j.tics.2011.11.001

Block, N. (2014). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences*, *18*(9), 445–447. https://doi.org/10.1016/j.tics.2014.05.007

Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 complex). *eNeuro, 3*(4). https://doi.org/10.1523/ENEURO.0039-16.2016

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, *116*, 220–251. https://doi.org/10.1037/a0014462

Bowers, J. S. (2017). Grandmother cells and localist representations: A review of current thinking. *Language, Cognition and Neuroscience, 32*, 257–273. https://doi.org/10.1080/23273798.2016.1267782

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience, 7*, 880–886. https://doi.org/10.1038/nn1278

Brincat, S. L., & Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron*, *49*, 17–24. https://doi.org/10.1016/j.neuron.2005.11.026

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience, 29*(44), 13992–14003. https://doi.org/10.1523/jneurosci.3577-09.2009

Brouwer, G. J., & Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *The Journal of Neuroscience, 33*(39), 15454–15465. https://doi.org/10.1523/JNEUROSCI.2472-13.2013

Buckthought, A., Kim, J., & Wilson, H. R. (2008). Hysteresis effects in stereopsis and binocular rivalry. *Vision Research*, *48*, 819–830. https://doi.org/10.1016/j.visres.2007.12.013

Cao, Y., Grossberg, S., & Markowitz, J. (2011). How does the brain rapidly learn and reorganize view-invariant and position-invariant object representations in the inferotemporal cortex? *Neural Networks, 24*(10), 1050–1061. https://doi.org/10.1016/j.neunet.2011.04.004

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing, 37*, 54–115.

Carpenter, G. A., & Grossberg, S. (2003). Adaptive resonance theory. In M.A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks, Second Edition* (pp. 87–90). MIT Press.

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks, 4*(6), 759–771. https://doi.org/10.1016/0893-6080(91)90056-B

Carpenter, G. A., Milenova, B. L., & Noeske, B. W. (1998). Distributed ARTMAP: A neurla network for fast distibuted supervised learning. *Neural Networks, 11*(5), 793–813. https://doi.org/10.1016/S0893-6080(98)00019-7

Chang, H. C., Grossberg, S., & Cao, Y. (2014). Where's Waldo? How perceptual, cognitive, and emotional brain processes cooperate during learning to categorize and find desired objects in a cluttered scene. *Frontiers in Integrative Neuroscience, 8*, 43. https://doi.org/10.3389/fnint.2014.00043

Chang, L., Bao, P., & Tsao, D. Y. (2017). The representation of colored objects in macaque color patches. *Nature Communications, 8*(1), 2064. https://doi.org/10.1038/s41467-017-01912-7

Chetverikov, A., Campana, G., & Kristjansson, A. (2017). Representing color ensembles. *Psychological Science*, *28*(10), 1–8. https://doi.org/10.1177/0956797617713787

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 1–73. https://doi.org/10.1017/S0140525X12000477

Clifford, A., Franklin, A., Holmes, A., Drivonikou, V. G., Ozgen, E., & Davies, I. R. (2012). Neural correlates of acquired color category effects. *Brain and cognition, 80*(1), 126–143. https://doi.org/10.1016/j.bandc.2012.04.011

Conway, B. R. (2009). Color vision, cones, and color-coding in the cortex. *Neuroscientist, 15*(3), 274–290. https://doi.org/10.1177/1073858408331369

Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized color modules in macaque extrastriate cortex. *Neuron, 56*(3), 560–573. https://doi.org/dmrm75

Conway, B. R., & Tsao, D. Y. (2009). Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 106*(42), 18034–18039. https://doi.org/10.1073/pnas.0810943106

Delk, J. L., & Fillenbaum, S. (1965). Differences in perceived color as a function of characteristic color. *The American Journal of Psychology, 78,* 290–293.

Deroy, O. (2013). Object-sensitivity versus cognitive penetrability of perception. *Philosophical Studies, 162,* 87–107. https://doi.org/10.1007/s11098-012-9989-1

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222. https://doi.org/bmcht5

Domijan, D., & Šetić, M. (2016). Resonant dynamics of grounded cognition: Explanation of behavioral and neuroimaging data using the ART neural network. *Frontiers in Psychology, 7*(139), 1–13. https://doi.org/10.3389/fpsyg.2016.00139

Duncan, J. (1980a). The demonstration of capacity limitation. *Cognitive Psychology, 12*(1), 75–96. https://doi.org/10.1016/0010-0285(80)90004-3

Duncan, J. (1980b). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87*(3), 272–300. https://doi.org/10.1037/0033-295X.87.3.272

Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017a). Variations in normal color vision. VI. Factors underlying individual differences in hue scaling and their implications for models of color appearance. *Vision Research, 141*, 51–65. https://doi.org/10.1016/j.visres.2016.12.006

Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017b). Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research, 141*, 66–75. https://doi.org/10.1016/j.visres.2016.12.007

van Es, D. M., Theeuwes, J., & Knapen, T. (2018). Spatial sampling in human visual cortex is modulated by both spatial and feature-based attention. *Elife, 7*, https://doi.org/10.7554/eLife.36928

Escobar, W. (2013). Quantized visual awareness. *Frontiers in Psychology, 4*(869). https://doi.org/10.3389/fpsyg.2013.00869

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18*(4), 500–549.

Estes, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

Fazl, A., Grossberg, S., & Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology, 58*(1), 1–48. https://doi.org/10.1016/j.cogpsych.2008.05.001

Ffytche, D. H., & Zeki, S. (2011). The primary visual cortex, and feedback to it, are not necessary for conscious vision. *Brain, 134*(Pt 1), 247–257. https://doi.org/10.1093/brain/awq305

Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science*, *25*, 38–46. https://doi.org/10.1177/0956797613485092

Firestone, C., & Scholl, B. J. (2015a). Can you experience top-down effects on perception? The case of race categories and perceived lightness. *Psychonomic Bulletin & Review*, *22*, 694–700. https://doi.org/10.3758/s13423-014-0711-5

Firestone, C., & Scholl, B. J. (2015b). Enhanced visual awareness for morality and pajamas? Perception vs. memory in "top-down" effects. *Cognition, 136,* 409–416. https://doi.org/10.1016/j.cognition.2014.10.014

Firestone, C., & Scholl, B. J. (2015c). When do ratings implicate perception vs. judgment? The "overgeneralization test" for top-down effects. *Visual Cognition*, *23*, 1217–1226. https://doi.org/10.1080/13506285.2016.1160171

Firestone, C., & Scholl, B. J. (2017). Seeing and thinking in studies of embodied "perception". *Perspectives on Psychological Science*, *12*(2), 341–343. https://doi.org/gh4g

Foley, N. C., Grossberg, S., & Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: Object cueing, useful-field-of-view, and crowding. *Cognitive Psychology, 65*(1), 77–117. https://doi.org/10.1016/j.cogpsych.2012.02.001

Forder, L., He, X., & Franklin, A. (2017). Colour categories are reflected in sensory stages of colour perception when stimulus issues are resolved. *PLoS ONE, 12*(5), e0178097. https://doi.org/10.1371/journal.pone.0178097

Foster, D. H. (2011). Color constancy. *Vision Research, 51*(7), 674–700. https://doi.org/10.1016/j.visres.2010.09.006

Francis, G. (2010). Modeling filling-in of afterimages. *Attention, Perception, & Psychophysics, 72*(1), 19–22. https://doi.org/10.3758/app.72.1.19

Francis, G. (2012). The same old New Look: Publication bias in a study of wishful seeing. *i-Perception, 3*(3), 176-178. https://doi.org/10.1068/i0519ic

Francis, G. (2019). *Hypothesis testing reconsidered* (Elements in perception). Cambridge University Press. https://doi.org/10.1017/9781108582995

Francis, G., & Ericson, J. (2004). Using afterimages to test neural mechanisms for perceptual filling-in. *Neural Networks, 17*, 737–752. https://doi.org/10.1016/j.neunet.2004.01.007

Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research, 34,* 1089–1104.

Francis, G., Manassi, M., & Herzog, M. H. (2017). Neural dynamics of grouping and segmentation explain properties of visual crowding. *Psychological Review*, *124*(4), 483–504. https://doi.org/10.1037/rev0000070.

Francis, G., & Thunell, E. (2019). Excess success in "Ray of hope: Hopelessness increases preferences for brighter lighting". *Collabra: Psychology*, *5*(1), 22. https://doi.org/10.1525/collabra.213

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences, 3*(4), 128–135. https://doi.org/10.1016/s1364-6613(99)01294-2

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience, 14*(5), 350–363. https://doi.org/10.1038/nrn3476

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science, 6*, 298–394. https://doi.org/10.1111/j.1467-9280.1995.tb00514.x

Goldstone, R. L., de Leeuw, J. R., & Landy, D. H. (2015). Fitting perception in and to cognition. *Cognition*, *135,* 24–29. https://doi.org/10.1016/j.cognition.2014.11.027

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87*(1), 1–51. https://doi.org/10.1037/0033-295X.87.1.1

Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and Cognition, 8,* 1–44. https://doi.org/10.1006/ccog.1998.0372

Grossberg, S. (2000). How hallucinations may arise from brain mechanisms of learning, attention, and volition. *Journal of the International Neuropsychological Society, 6*(5), 583–592. https://doi.org/10.1017/s135561770065508x

Grossberg, S. (2003). How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. *Behavioral and Cognitive Neuroscience Reviews*, *2*, 47–76. https://doi.org/10.1177/1534582303002001003

Grossberg, S. (2013). Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks, 37*, 1–47. https://doi.org/10.1016/j.neunet.2012.09.017

Grossberg, S. (2017a). Grandmother cohorts: Multiple-scale brain compression dynamics during learning of object and sequence categories. *Language, Cognition and Neuroscience, 32*, 295–315. https://doi.org/10.1080/23273798.2016.1232838

Grossberg, S. (2017b). Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Networks, 87*, 38–95. https://doi.org/org/10.1016/j.neunet.2016.11.003

Grossberg, S., Hwang, S., & Mingolla, E. (2002). Thalamocortical dynamics of the McCollough effect: Boundary-surface alignment through perceptual learning. *Vision Research, 42*(10), 1259–1286. https://doi.org/10.1016/s0042-6989(02)00055-x

Hansen, T., & Gegenfurtner, K. R. (2009). Independence of color and luminance edges in natural scenes. *Visual Neuroscience, 26*(1), 35–49. https://doi.org/czxtfh

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367–1368. https://doi.org/10.1038/nn1794

Harris, K. D., & Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience, 12*(9), 509–523. https://doi.org/10.1038/nrn3084

Haykin, S. (2009). *Neural networks and learning machines*. *Third edition*. Pearson Education.

He, X., Witzel, C., Forder, L., Clifford, A., & Franklin, A. (2014). Color categories only affect post-perceptual processes when same- and different-category colors are equally discriminable. *Journal of the Optical Society of America A, 31*(4), A322–A331. https://doi.org/10.1364/JOSAA.31.00A322

Heinke, D., & Humphreys, G. W. (2003). Attention, spatial representation, and visual neglect: Simulating emergent attention and spatial memory in the selective attention for identification model (SAIM). *Psychological Review, 110*(1), 29–87. https://doi.org/10.1037/0033-295x.110.1.29

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804. https://doi.org/10.1016/s0896-6273(02)01091-7

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hohwy, J. (2017). Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Consciousness and Cognition, 47*, 75–85. https://doi.org/10.1016/j.concog.2016.09.004

Hong, S. W., & Tong, F. (2017). Neural representation of form-contingent color filling-in in the early visual cortex. *Journal of Vision, 17*(13), 10. https://doi.org/10.1167/17.13.10

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114*(3), 599–631. https://doi.org/10.1037/0033-295x.114.3.599

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, 317, 823–825. https://doi.org/10.1126/science.1143515

Hurlbert, A. (1996). Colour vision: Putting it in context. *Current Biology, 6*(11), 1381–1384. https://doi.org/10.1016/S0960-9822(96)00736-1

Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. S. (2017). Perception of ensemble statistics require attention. *Consciousness and Cognition*, *48,* 149–160. https://doi.org/10.1016/j.concog.2016.11.007

Julesz, B. (1974). Cooperative phenomena in binocular depth perception. *American Scientist*, *62,* 32–43.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research, 46*(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America, 114*(13), 3521–3526. https://doi.org/10.1073/pnas.1611835114

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience, 17*(5), 307–321. https://doi.org/10.1038/nrn.2016.22

Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nature Reviews Neuroscience, 7*(3), 220–231. https://doi.org/10.1038/nrn1869

Kuriki, I. (2004). Testing the possibility of average-color perception from multi-colored patterns. *Optical Review*, *11*, 249–257. https://doi.org/10.1007/s10043-004-0249-2

Lafer-Sousa, R., & Conway, B. R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nature Neuroscience, 16*(12), 1870–1878. https://doi.org/10.1038/nn.3555

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227–260. https://doi.org/10.1037/0033-295x.107.2.227

Lamme, V. A. F. (2001). Blindsight: The role of feedforward and feedback corticocortical connections. *Acta psychologica, 107*(1–3), 209–228. https://doi.org/10.1016/s0001-6918(01)00020-8

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*(1), 12–18. https://doi.org/10.1016/s1364-6613(02)00013-x

Lamme, V. A. F. (2018). Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 373*(1755). https://doi.org/10.1098/rstb.2017.0344

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27*, 889–935. https://doi.org/10.1068/p270889

Leonard, C. J., Balestreri, A., & Luck, S. J. (2015). Interactions between space-based and feature-based attention. *Journal of Experimental Psychology. Human Perception and Performance, 41*(1), 11–16. https://doi.org/10.1037/xhp0000011

Leopold, D. A. (2012). Primary visual cortex: awareness and blindsight. *Annual Review of Neuroscience, 35*, 91–109. https://doi.org/10.1146/annurev-neuro-062111-150356

Li, M., Liu, F., Juusola, M., & Tang, S. (2014). Perceptual color map in macaque visual area V4. *The Journal of Neuroscience, 34*(1), 202–217. https://doi.org/f5m4b2

Lim, H., Wang, Y., Xiao, Y., Hu, M., & Felleman, D. J. (2009). Organization of hue selectivity in macaque V2 thin stripes. *Journal of Neurophysiology, 102*(5), 2603–2615. https://doi.org/10.1152/jn.91255.2008

Linhares, J. M., Pinto, P. D., & Nascimento, S. M. (2008). The number of discernible colors in natural scenes. *Journal of the Optical Society of America. A, Optics, image science, and vision, 25*(12), 2918–2924. https://doi.org/10.1364/josaa.25.002918

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Second edition. MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label feedback hypothesis. *Frontiers in Cognition, 3(54).* https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G. (2015a). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology, 6*(4), 547–569. https://doi.org/10.1007/s13164-015-0253-4

Lupyan, G. (2015b). Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*. *161,* 117–130. https://doi.org/ghh8c7

Lupyan, G. (2017a). Changing what you see by changing what you know: The role of attention. *Frontiers in Psychology, 8*(553). https://doi.org/10.3389/fpsyg.2017.00553

Lupyan, G. (2017b). How reliable is perception? *Philosophical Topics, 45*(1), 81–106. https://doi.org/10.17605/OSF.IO/R7SJJ

Macknik, S. L., & Martinez-Conde, S. (2009). The role of feedback in visual attention and awareness. In M. S. Gazzaniga (Ed.), *The cognitive neuroscience* (pp. 1165–1179). MIT Press.

Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research, 84*(1), 24–62. https://doi.org/10.1111/j.1933-1592.2010.00481.x

Macpherson, F. (2017). The relationship between cognitive penetration and predictive coding. *Consciousness and Cognition, 47,* 6–16. https://doi.org/10.1016/j.concog.2016.04.001

von der Malsburg, C. (1999). The what and why of binding: the modeler's perspective. *Neuron, 24*(1), 95–104. https://doi.org/10.1016/s0896-6273(00)80825-9

Marić, M., & Domijan, D. (2018). A neurodynamic model of feature-based spatial selection. *Frontiers in Psychology*, *9*(417), 1-22. https://doi.org/10.3389/fpsyg.2018.00417

Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science, 342*(6158), 1238406. https://doi.org/10.1126/science.1238406

Markov, N. T., & Kennedy, H. (2013). The importance of being hierarchical. *Current Opinion in Neurobiology*, *23*(2), 187–194. https://doi.org/10.1016/j.conb.2012.12.008

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K., Kennedy, H. (2014). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, *522*(1), 225–259. https://doi.org/10.1002/cne.23458

Masuyama, N., Loo, C. K., & Dawood, F. (2018). Kernel Bayesian ART and ARTMAP. *Neural Networks*, *98,* 76–86. https://doi.org/10.1016/j.neunet.2017.11.003.

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, *15*(4), 6. https://doi.org/10.1167/15.4.6

Maule, J., & Franklin, A. (2016). Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism. *Journal of the Optical Society of America A:*

Optics, Image Science, and Vision, 33, A22–A29. https://doi.org/10.1364/JOSAA.33.000A22

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*(4), 287–330. https://doi.org/10.1037/0033-295X.86.4.287

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation: Volume 24* (pp. 109–165). San Diego: Academic Press. https://doi.org/10.1016/S0079-7421(08)60536-8

Medin, D. L., & Schafer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J. M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron, 89*(2), 384–397. https://doi.org/10.1016/j.neuron.2015.12.018

Newen, A., & Vetter, P. (2017). Why cognitive penetration of our perceptual experience is still the most plausible account. *Consciousness and Cognition, 47,* 26–37. https://doi.org/10.1016/j.concog.2016.09.005

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology*: *General, 115,* 39–57. https://doi.org/10.1037//0096-3445.115.1.39

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. Wills (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge University Press.

O'Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., Jr., & Bar, M. (2017). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition, 47*, 63–74. https://doi.org/10.1016/j.concog.2016.05.003

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision, 8*(5), 13.11–16. https://doi.org/10.1167/8.5.13

Olman, C. A. (2015). What insights can fMRI offer into the structure and function of mid-tier visual areas? *Visual Neuroscience, 32*: E015. https://doi.org/ghh5wn

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13*(11), 4700–4719. https://doi.org/gg2gvn

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1995). A multiscale dynamic routing circuit for forming size- and position-invariant object representations. *Journal of Computational Neuroscience, 2*(1), 45–62. https://doi.org/10.1007/bf00962707

Ostergaard, M. (2018). Phenomenal consciousness and cognitive access. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 373*(1755): 20170353. https://doi.org/10.1098/rstb.2017.0353

Papale, P., Leo, A., Cecchetti, L., Handjaras, G., Kay, K. N., Pietrini, P., & Ricciardi, E. (2018). Foreground-background segmentation revealed during natural image viewing. *eNeuro, 5*(3). https://doi.org/10.1523/ENEURO.0075-18.2018

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292*(5516), 510–512. https://doi.org/10.1126/science.1057099

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, *82*, 2490–2502. https://doi.org/10.1152/jn.1999.82.5.2490

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of Neurophysiology*, *86*, 2505–2519.

Pfülb B., Gepperth A., Abdullah S., Kilian A. (2018). Catastrophic forgetting: Still a problem for DNNs. In V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis (Eds.) *Artificial Neural Networks and Machine Learning – ICANN 2018. Lecture Notes in Computer Science, Vol. 11139* (pp. 487–497). Springer: Cham. https://doi.org/10.1007/978-3-030-01418-6_48

Poltoratski, S., & Tong, F. (2014). Hysteresis in the dynamics perception of scenes and objects. *Journal of Experimental Psychology: General*, *143*(5), 1875–1892. https://doi.org/10.1037/a0037365

Prinz, J. J. (2000). A neurofunctional theory of visual consciousness. *Consciousness and Cognition, 9*, 243–259. https://doi.org/10.1006/ccog.2000.0442

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*, 341–365. https://doi.org/10.1017/s0140525x99002022

Raftopoulos, A. (2001). Is perception informationally encapsulated? The issue of the theory-ladenness of perception. *Cognitive Science, 25*, 423–451. https://doi.org/10.1016/S0364-0213(01)00042-8

Raftopoulos, A. (2009). *Cognition and perception: How do psychology and neural science inform philosophy?* MIT Press.

Raftopoulos, A. (2014). The cognitive impenetrability of the content of early vision is a necessary and sufficient condition for purely nonconceptual content. *Philosophical Psychology, 27*(5), 601–620. https://doi.org/10.1080/09515089.2012.729486

Raftopoulos, A., & Zeimbekis, J. (2015). The cognitive penetrability of perception: An overview. In J. Zeimbekis & A. Raftopoulos (Eds.), *The Cognitive Penetrability of Perception: New Perspectives*. Oxford University Press.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience, 27,* 611–647. https://doi.org/cv7ndm

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19*(5), 1736–1753. https://doi.org/10.1523/JNEUROSCI.19-05-01736.1999

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., & Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron, 74*(1), 12–29. https://doi.org/10.1016/j.neuron.2012.03.011

Sasaki, Y., & Watanabe, T. (2004). The primary visual cortex fills in color. *Proceedings of the National Academy of Sciences, 101*(52), 18251–18256. https://doi.org/fc3c5c

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science*, *15*, 720–728. https://doi.org/10.1111/j.0956-7976.2004.00748.x

Seymour, K. J., Williams, M. A., & Rich, A. N. (2016). The representation of color across the human visual cortex: Distinguishing chromatic signals contributing to object form versus surface color. *Cerebral Cortex, 26*(5), 1997–2005. https://doi.org/10.1093/cercor/bhv021

Silvanto, J. (2015). Why is "blindsight" blind? A new perspective on primary visual cortex, recurrent activity and visual awareness. *Consciousness and Cognition, 32,* 15–32. https://doi.org/10.1016/j.concog.2014.08.001

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 360*(1458), 1329–1346. https://doi.org/10.1098/rstb.2005.1633

Song, H., Vonasch, A. J., Meier, B. P., & Bargh, J. A. (2012). Brighten up: Smiles facilitate perceptual judgement of facial lightness. *Journal of Experimental Social Psychology, 48,* 450–452. https://doi.org/10.1016/j.jesp.2011.10.003

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9): 403–409. http://dx.doi.org/10.1016/j.tics.2009.06.003

Summerfield, C., & Egner, T. (2014). Attention and decision making. In A. C. Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (pp. 837–864). Oxford University Press.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, 106*(11), 4567–4570. https://doi.org/b4783s

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522. https://doi.org/10.1038/381520a0

Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience, 4*(3), 219–229. https://doi.org/10.1038/nrn1055

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences, 24*(5), 295–300. https://doi.org/10.1016/s0166-2236(00)01814-2

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. W. J. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549–586). MIT Press.

Valenti, J. J., & Firestone, C. (2019). Finding the "odd one out": Memory color effects and the logic of appearance. *Cognition, 191*, 103934. https://doi.org/gh4f

Van Dromme, I. C., Premereur, E., Verhoef, B. E., Vanduffel, W., & Janssen, P. (2016). Posterior parietal cortex drives inferotemporal activations during three-dimensional object vision. *PLoS Biology, 14*(4), e1002445. https://doi.org/f8wqvc

Vandenbroucke, A. R. E., Fahrenfort, J. J., Meuwese, J. D. I., Scholte, H. S., & Lamme, V. A. F. (2016). Prior knowledge about objects determines neural color representation in human visual cortex. *Cerebral Cortex, 26*(4), 1401–1408. https://doi.org/f8h9cn

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, 30*(6), 655–668. https://doi.org/bzm9gn

Velez, R., & Clune, J. (2017). Diffusion-based neuromodulation can eliminate catastrophic forgetting in simple neural networks. *PloS ONE, 12*(11), e0187736. https://doi.org/10.1371/journal.pone.0187736

Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition, 27,* 62–75. https://doi.org/10.1016/j.concog.2014.04.007

Vigdor, B., & Lerner, B. (2007). The Bayesian ARTMAP. *IEEE Transactions on Neural Networks, 18*(6), 1628–1644. https://doi.org/10.1109/tnn.2007.900234

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all-or-none samples. *Journal of Experimental Psychology: General*, *138,* 546–560. https://doi.org/10.1037/a0017352

Waldrop, M. M. (2019). What are the limits of deep learning? *Proceedings of the National Academy of Sciences of the United States of America*, *116* (4), 1074–1077. https://doi.org/10.1073/pnas.1821594116

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395–1407. https://doi.org/10.1016/j.neunet.2006.10.001

Ward, E. J. (2018). Downgraded phenomenology: How conscious overflow lost its richness. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *373*(1755): 20170355. https://doi.org/10.1098/rstb.2017.0355

Williams, D., & Sekuler, G. B. R. (1986). Hysteresis in the perception of motion direction as evidence for neural cooperativity. *Nature*, *324,* 253–255. https://doi.org/10.1038/324253a0

Williamson, J. R. (1996). Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps. *Neural Networks, 9*(5), 881–897. https://doi.org/10.1016/0893-6080(95)00115-8

Winawer, J., & Witthoft, N. (2015). Human V4 and ventral occipital retinotopic maps. *Visual Neuroscience, 32*, E020. https://doi.org/10.1017/s0952523815000176

Witzel, C. (2016). An easy way to show memory color effects. *i-Perception, 7*(5), 2041669516663751. https://doi.org/10.1177/2041669516663751

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object knowledge modulates colour appearance. *i-Perception, 2*(1), 13–49. https://doi.org/10.1068/i0396

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics, 72*(8), 2031–2046. https://doi.org/10.3758/app.72.8.2031

Xiao, Y., Wang, Y., & Felleman, D. J. (2003). A spatially organized representation of colour in macaque cortical area V2. *Nature*, *421*(6922), 535–539. https://doi.org/10.1038/nature01372

Zaidi, Q., Marshall, J., Thoen, H., & Conway, B. R. (2014). Evolution of neural computations: Mantis shrimp and human color decoding. *i-Perception*, *5*(6), 492–496. https://doi.org/10.1068/i0662sas

Zeimbekis, J. (2013). Color and cognitive penetrability. *Philosophical Studies*, *165*, 167–175. https://doi.org/10.1007/s11098-012-9928-1

APPENDIX E

Supplemental Materials

This appendix contains supplemental materials in the form of Open Science Framework links that generate the MATLAB code to reproduce all results reported in the doctoral thesis. The Matlab code for the papers listed in Appendix A, B, and D is available on:

- https://osf.io/5uf6g/
- https://osf.io/9h7ag/
- https://osf.io/zphyq/, respectively.