

2009-08-12

# Cortical Dynamics of Contextually-Cued Attentive Visual Learning and Search: Spatial and Object Evidence Accumulation

---

<https://hdl.handle.net/2144/1974>

*Downloaded from DSpace Repository, DSpace Institution's institutional repository*

# **Cortical dynamics of contextually-cued attentive visual learning and search: Spatial and object evidence accumulation**

**Tsung-Ren Huang and Stephen Grossberg<sup>\*</sup>**

Department of Cognitive and Neural Systems  
Center for Adaptive Systems  
and  
Center of Excellence for Learning in Education, Science, and Technology  
Boston University  
677 Beacon Street  
Boston, MA 02215  
Phone: 617-353-7858  
Fax: 617-353-7755  
E-mail: [tren@cns.bu.edu](mailto:tren@cns.bu.edu), [steve@cns.bu.edu](mailto:steve@cns.bu.edu)

Corresponding Author: Stephen Grossberg

**Technical Report CAS/CNS-TR-2009-010**

Submitted: August 12, 2009

**Keywords:** spatial contextual cueing; object contextual cueing; spatial attention; object attention; saliency map; visual search; scene perception; scene memory; implicit learning; prefrontal cortex; medial temporal lobe

---

TRH and SG are supported in part by CELEST, an NSF Science of Learning Center (SBE-0354378), and the SyNAPSE program of DARPA (HR0011-09-3-0001, HR0011-09-C-0011)

Copyright © 2009

Permission to copy without fee all or part of this material is granted provided that: 1. The copies are not made or distributed for direct commercial advantage; 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and / or special permission.

## Abstract

How do humans use predictive contextual information to facilitate visual search? How are consistently paired scenic objects and positions learned and used to more efficiently guide search in familiar scenes? For example, a certain combination of objects can define a context for a kitchen and trigger a more efficient search for a typical object, such as a sink, in that context. A neural model, ARTSCENE Search, is developed to illustrate the neural mechanisms of such memory-based contextual learning and guidance, and to explain challenging behavioral data on positive/negative, spatial/object, and local/distant global cueing effects during visual search. The model proposes how global scene layout at a first glance rapidly forms a hypothesis about the target location. This hypothesis is then incrementally refined by enhancing target-like objects in space as a scene is scanned with saccadic eye movements. The model clarifies the functional roles of neuroanatomical, neurophysiological, and neuroimaging data in visual search for a desired goal object. In particular, the model simulates the interactive dynamics of spatial and object contextual cueing in the cortical What and Where streams starting from early visual areas through medial temporal lobe to prefrontal cortex. After learning, model dorsolateral prefrontal cortical cells (area 46) prime possible target locations in posterior parietal cortex based on goal-modulated percepts of spatial scene gist represented in parahippocampal cortex, whereas model ventral prefrontal cortical cells (area 47/12) prime possible target object representations in inferior temporal cortex based on the history of viewed objects represented in perirhinal cortex. The model hereby predicts how the cortical What and Where streams cooperate during scene perception, learning, and memory to accumulate evidence over time to drive efficient visual search of familiar scenes.

## 1. Introduction: Context- and Goal-Dependent Scene Understanding

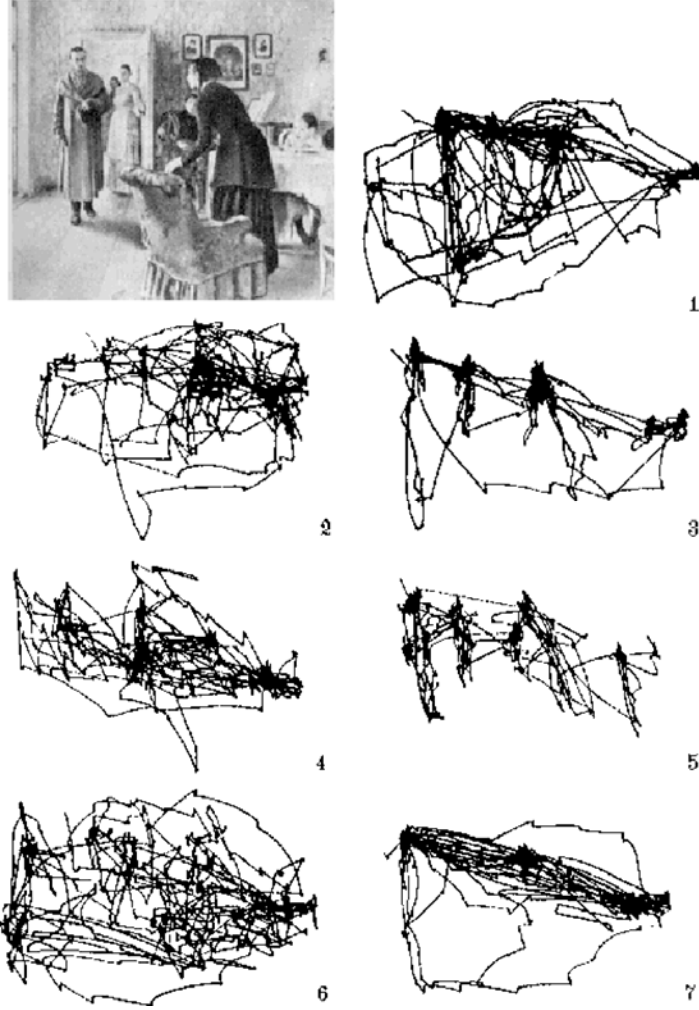
We make thousands of eye movements every day. Our visual attention and eye movements explore scenes without any goals in mind. Just as often, however, we search for valued targets embedded in complex visual scenes. Common examples include finding a friend in a crowd or locating a menu board in a café. To search efficiently, people prioritize visual attention using the knowledge of what to expect and where to look. Such knowledge comes either from exogenous cues, such as visual or verbal hints of the target, or from endogenous memories of spatial or object regularities in a scene.

Scene gist, a rapid yet crude representation of a scene, helps human observers to deploy visual attention prior to eye movements. Behavioral data have shown that human observers process visual information in a global-to-local and coarse-to-fine manner (Navon, 1977; Schyns & Oliva, 1994). Just by the first glance of a novel image in ~ 200-300ms, people are able to recognize the basic-level scene identity (Potter, 1976; Tversky & Hemenway, 1983), and grasp surface properties (Oliva & Schyns, 2000; Rousselet, Joubert, & Fabre-Thorpe, 2005), spatial structures (Biederman, Rabinowitz, Glass, & Stacy, 1974; Sanocki, 2003), and meanings (Potter, 1975; Potter, Staub, & O' Connor, 2004) without parsing individual objects in the scene. Such expeditious comprehension of a scene also provides contextual guidance on where a search target may be located (Torralba, Oliva, Castelhano, Henderson, 2006).

Percepts of scene gist from a single fixation are often only the first-order approximation to scene understanding. Evidence accumulation over time is also recognized as a fundamental computation for primate visual perception and cognition (Jonides, Irwin, & Yantis, 1982; Irwin, 1991; Gold & Shadlen, 2007; Heekeren, Marrett, & Ungerleider, 2008; Grossberg & Pilly, 2008). Recent neural models clarify how successive spatial attention shifts and eye movements can offer a higher-order, progressively developing understanding of scenes (Grossberg & Huang, 2009) and the objects within them (Fazl, Grossberg, & Mingolla, 2009).

Since visual attention can be allocated volitionally to objects or regions of interest, scan paths and fixation hotspots of eye movements are not solely dependent on structural statistics of a scene, but also reflect internal drives or task-dependent goals. For instance, when geologists walk into a desert, their attention may be attracted to massive brownish formations for their field studies. However, if they are desperately thirsty when arriving at the same place with the same view, their eyes may first check bluish spots in the field, under the hope of seeing an oasis. Indeed, Yarbus (1967) has provided a classical example of such goal-dependent scene search by recording eye movements for the same picture under different task instructions (Figure 1).

As a direct result of attention-modulated scene percepts, the memory of a scene is not a verbatim copy of the external world, but is rather a modulated map whose principal components are attentionally salient textures or objects (Kensinger, Garoff-Eaton, & Schacter, 2007). It also follows that the attentional saliency of a local entity in a scene can be contributed from bottom-up perceptual factors as well as top-down cognitive (Leber & Egeth, 2006; Chen & Zelinsky, 2006) or emotional primes (Öhman, Flykt, & Esteves, 2001; Armony & Dolan, 2002).

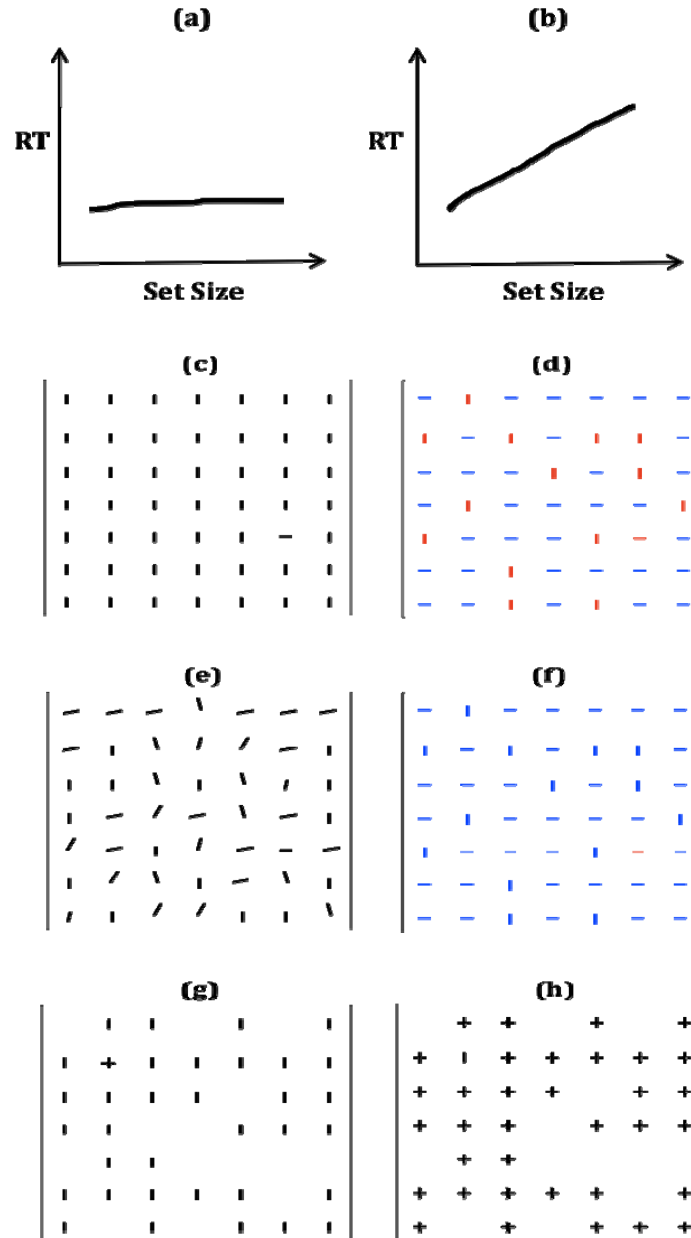


**Figure 1.** Eye-scan paths and fixation hotspots of an observer varied given different task instructions. Conditions in each panel are (1) free viewing, (2) estimating the wealth of the family, (3) judging their ages, (4) guessing what they had been doing before the arrival of the unexpected visitor, (5) remembering the clothes worn by the people, (6) memorizing the location of the people and objects in the painting, and (7), estimating how long the unexpected visitor had been away from the family. (Reprint with permission from Yarbus, 1967).

The challenges of a complete visual scene understanding theory are to clarify how exogenous and endogenous attention dynamically organize scene perception and memory, and how the neural dynamics of evidence accumulation incrementally deepens awareness and knowledge of a scene in the course of spatial attention shifts and exploratory eye movement scanning sequences. The ARTSCENE model (Grossberg & Huang, 2009) simulated how spatial attention can regulate category learning and recognition of scenic textures from global to local scales to advance scene identification over time. Here the ARTSCENE Search model is developed to illustrate how global-to-local evidence accumulation, combined with learned contextual information from multiple objects and positions in a scene, can quantitatively explain a large set of visual search data, with a special emphasis on attention-modulated memory-based contextual effects.

## 2. A Brief Review of Psychophysical and Modeling Studies of Visual Search

Visual search is a task that requires active eye scans to locate target features or objects among distractors in a visual environment. In laboratory settings, a target is predefined, either verbally or by visual exposure, and a search display is usually a two-dimensional photographic or naturalistic scene, or simply composed of colored bars, English letters, or basic geometric shapes like circles and triangles.



**Figure 2.** Summary of basic search properties: (a) Zero slope in efficient search. (b) Nonzero slope in inefficient search. (c) Efficient feature search for a horizontal bar. (d) Inefficient conjunction search for a red horizontal bar. (e) Inefficient feature search for a horizontal bar. (f) Efficient conjunction search for a red horizontal bar. (g) Efficient search for a cross. (h) Inefficient search for a vertical bar.

To quantify search performance, reaction time (RT), the elapsed time between presentation of a search screen and discovery of a task target, is usually recorded in psychophysical studies. This measure reveals both quantitative and qualitative differences between experimental conditions. RT is often evaluated as a linear function of set size (i.e., the total number of items in a search display). The corresponding slope and intercept characterize search efficiency and time for perceptual processing plus response selection, respectively. When a target is defined by a distinctive attribute such as color, size, orientation, or shape, search is often efficient and a target pops out from the background for all set sizes (Figure 2c), producing a zero search slope (Figure 2a). In contrast, when a target is absent or defined by a conjunction of basic attributes that are also shared by distractors (Figure 2d), search is often inefficient and RT increases in proportion to set size, producing a non-zero search slope (Figure 2b).

Based on the seeming dichotomy of efficient feature search versus inefficient conjunction search, Treisman & Gelade (1980) proposed a two-stage visual attention model named Feature Integration Theory (FIT) in which primitive features are processed within their own feature map in a rapid, pre-attentive, and parallel manner, followed by the second slow stage where serial deployment of spatial attention binds features into an object at the attended location for object recognition or further processing. Segregated feature maps in the first stage of FIT were supported by the finding that size, motion, and orientation made additive and independent contributions to the search slope of a double conjunction search (Treisman & Sato, 1990). Attentive feature binding in the second stage of FIT was supported by the percept of illusory conjunctions of features from different items in the same display, especially when attention is overloaded or diverted in a rapid search task (Treisman & Schmidt, 1982).

The dichotomy of efficient versus inefficient search based on slopes was later shown to be inadequate. A continuum of flat to steep slopes can be obtained by varying saliency factors (Wolfe, Cave, & Franzel, 1989; Wolfe, 1998). In particular, search efficiency increases with decreased similarity of targets to distractors and increased similarity between distractors (Duncan & Humphreys, 1989). In other words, a feature search can be inefficient (Figure 2e) and a conjunction search can be efficient (Figure 2f), all depending on the degree to which a target can be distinguished from distractors.

The newly observed efficient conjunctive searches can be explained by either top-down or bottom-up factors. In the class of top-down models, parallel enhancement of target features (Guided Search, Wolfe et al, 1989; Wolfe, 1994) or suppression of non-target features (Revised FIT, Treisman & Sato, 1990) were introduced to the FIT architecture to bias spatial selection toward target locations and thus bypass the need for serial conjunctions at distractor locations. Here, the priori knowledge of the target can be exogenously specified or endogenously acquired from scene memory (Contextual Cueing, Chun & Jiang, 1998; Chun & Jiang, 1999). In the class of bottom-up models (Attentional Engagement Theory, Duncan & Humphreys, 1989; SERR, Humphreys & Müller, 1993; SOS, Grossberg, Mingolla, & Ross, 1994; CTVA, Logan, 1996), perceptual grouping based on featural similarity and spatial proximity between objects dynamically organizes items for parallel processing, and effectively reduces the set size for serial operations such as attention reallocation. In principle, these two classes of models do not contradict with each other. Perceptual grouping can be realized within each feature map as part of parallel processing in FIT. Indeed, in the Spatial and Object Search (SOS) model of Grossberg, Mingolla, & Ross (1994), bottom-up grouping and surface color factors set the stage for the serial allocation of top-down spatial and object attention.

In this regard, a parallel line of development of the FIT framework concerns how spatial selection and attention shifts can be carried out with plausible brain mechanisms. Koch & Ullman (1985) proposed that individual feature maps of color, orientation, motion, disparity, etc. are normalized and integrated into a scalar saliency map, which represents the overall conspicuity of an object in space as a priority measure for attentional selection. Specifically, a location of maximum saliency in the map is selected first through winner-take-all competition with other locations, followed by suppression of activity at the selected location to implement inhibition of return (Posner, M., I., & Cohen, 1984; also see reviews in Klein, 2000) whereby attention can disengage from the winner location and continue a new selection cycle (see also Grossberg (1978) and Grossberg & Kuperstein (1986) for examples of saliency choice and inhibition of return). As a result, a saliency map (SM) is scanned in order of decreasing saliency by the focus of attention.

Computational refinements of the saliency model (Niebur & Koch, 1996; Itti & Koch, 2001) were applied to simple laboratory stimuli as well as complex natural scenes to compare with human data. In these algorithms, feature saliency is derived from multi-scale center-surround competition among locations in each feature map. Such center-surround mechanism highlights a locally distinctive feature, and may underlie search RT asymmetries (Treisman & Gormican, 1988; Li, 2002) when a target swaps identity with distractors (Figure 2g & 2h). In general, center-surround competition is ubiquitous in neural systems (von Békésy, 1967) and has been shown in neural models to be fundamentally important to visual perception and perceptual decision making (Grossberg, 1973, 1980, 1988; Grossberg & Pilly, 2008).

Although early studies focused more on perception-based attentional factors, recent models started to explicitly address learning and memory issues in visual search. For example, Grossberg, Mingolla, & Ross (1994) propose how target categories learned in an Adaptive Resonance Theory framework could focus learned target categories and object attention to influence visual search. In a related approach, Navalpakkam & Itti (2005) proposed that the learned feature memory of a target can be used for attentional biasing and object recognition. Torralba, Oliva, Castelhano, & Henderson (2006) used a Bayesian approach to show how learned spatial regularities of an object in a scene can guide spatial attention toward possible target zones (see Figure 3e). Backhaus, Heinke, & Humphreys (2005) used an associative memory to explain basic spatial cueing effects reported by Chun & Jiang (1998). Brady & Chun (2007) simulated the locality of spatial cueing effects (Olson & Chun, 200) by exponentially weighting input patterns surrounding a target.

None of these memory-based models provides a unified framework for learned evidence accumulation that incrementally integrates all available spatial and object constraint to limit the search space. Eye movements in most search models merely function for target checking, and the only dynamics in the basic saliency model is inhibition of return at selected locations. ARTSCENE Search, on the other hand, emphasizes that each fixation also contributes extra evidence to target location and identity, revising saccadic plans on the fly. Moreover, unlike other models, ARTSCENE Search clarifies neural data about how multiple cortical areas cooperate to use object and spatial contextual information to guide efficient visual search, learning, and recognition.

ARTSCENE Search illustrates how humans can direct spatial and feature-based attention to parse and encode a scene into memory and how scene memory is then recollected to facilitate visual search in a familiar environment. In the model, the search strategy is a global-to-local process whereby spatial contextual cueing (Chun & Jiang, 1998) is induced early based on the



spatial gist of a scene, whereas object contextual cueing (Chun & Jiang, 1999) is gradually developed based on the identities of non-target objects after each eye fixation. In ARTSCENE Search, context is a guidepost to its paired target. Specifically, memory-based contextual guidance is achieved by a series of associative votes from context objects/locations to a target object/location, where association strength is commensurate with co-occurrence frequency and attentional valence of both the search target/location and a context object/location. The attentional valence is defined here as the degree to which an object attracts attention in response to both bottom-up and top-down factors. Taken together, these design constraints allow ARTSCENE Search to explain and quantitatively simulate a wide range of phenomena in memory-based visual search, which are reviewed in the next section.

### **3. Contextual Cueing Effects in Visual Search**

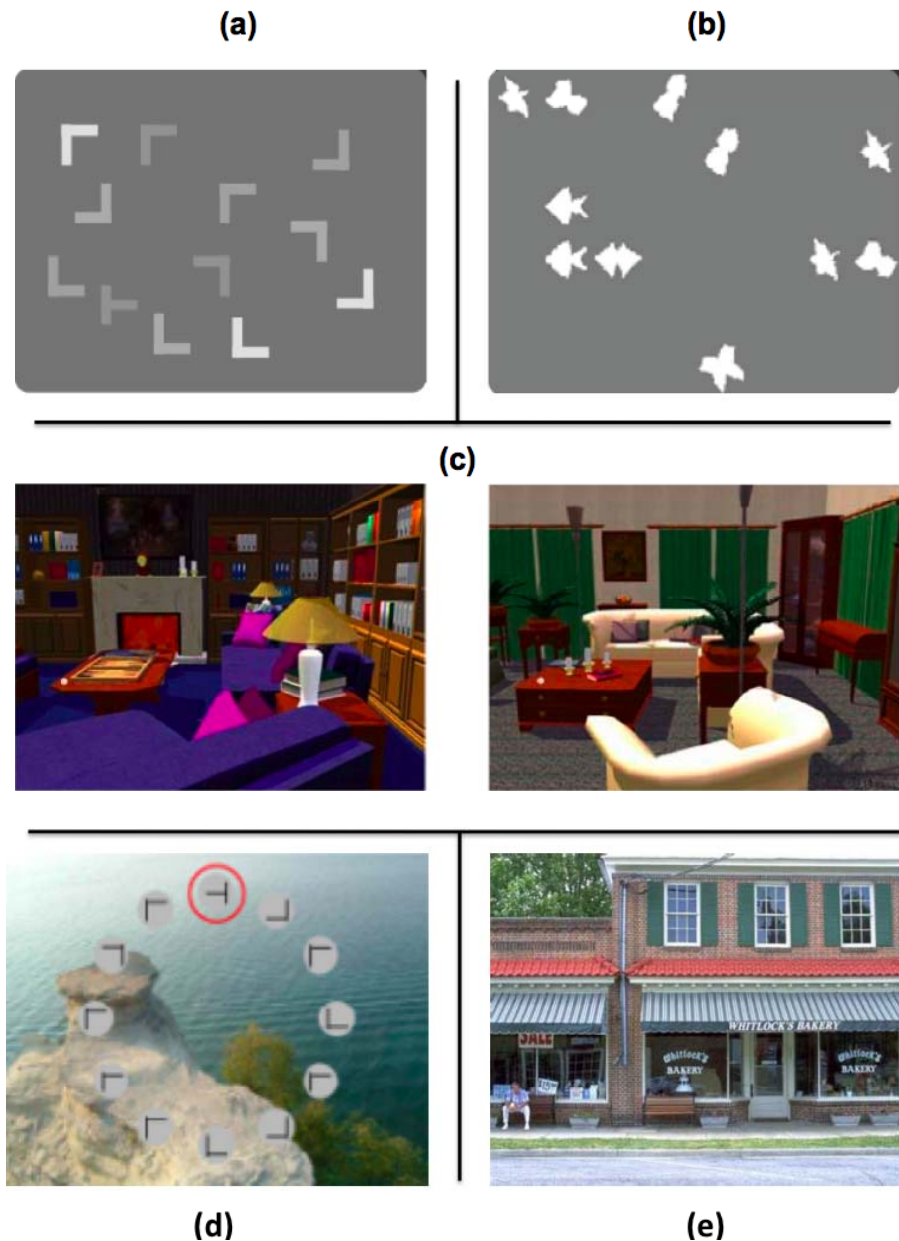
Efficient visual search exploits the memory of spatial and object regularities of a scene. For example, when we are looking for a friend in a beach picture, we direct our eyes right away to the bottom sand rather than the top sky. Such knowledge about the spatial layout of a scene is named spatial contextual cueing (Chun & Jiang, 1998). However, such spatial information is not always available in a new environment. For instance, when we are seeking beverages in a friend's refrigerator for the very first time, we may not even know where the kitchen is situated until seeing some related objects such as an oven or a microwave. In this scenario, object contextual cueing (Chun & Jiang, 1999) alludes the existence of a refrigerator in the kitchen context.

Psychophysically, contextual cueing effects are defined as the RT difference in visual search between familiar and novel scenes. Compared with semantic object cueing, spatial cueing is investigated much more thoroughly. Studies of spatial cueing often have discrete objects arranged in an invisible grid, and a target is either paired with a novel or invariant background of non-target objects across training blocks (see Figure 3 for stimulus samples). Lately, spatial cueing effects are also shown with nature (Brockmole & Henderson, 2006; Torralba, Oliva, Castelano, & Henderson, 2006) or naturalistic (Brockmole, Castelano, & Henderson, 2006) scenes. Although learning of repeated contexts is often reported to be implicit without subjects' awareness of the target-context covariation in letter displays (Chun & Jiang, 1998), cueing effects can also be obtained from explicit context learning of real scenes (Brockmole & Henderson, 2006). Moreover, memory of spatial context persists for at least one week once acquired (Chun & Jiang, 2003; Jiang, Song, & Rigas, 2005).

Spatial context, rapidly apprehended as the spatial gist of a scene, guides further allocation of focal attention. Such spatial cueing can occur in 200ms (Chun & Jiang, 1998, Experiment 5), which is less than the average fixation time (~300ms) (Potter, 1975). In other words, a single glance of a familiar scene suffices to frame a more efficient search. Consistent with the global and coarse nature of gist processing, spatial cueing can occur in a familiar global layout composed of locally jittered items (Chun & Jiang, 1998, Experiment 6), or substituted component objects (Chun & Jiang, 1998, Experiment 2). More importantly, spatial cueing expedites visual search by reducing the number of saccades during ineffective search (Tseng & Li, 2004), reflecting a strategy change of spatial selection in a repeated search display.

Data from Kunar, Flusberg, Horowitz, & Wolfe (2007) showed a weak relationship between attentional guidance and contextual cueing, and indicated perceptual processing and response selection to be the main factors for reducing search RT in spatial cueing experiments. However, unlike other spatial cueing studies, they enlarged items with increasing eccentricity with respect to the center of the search display, while allowing subjects to move eyes freely

during visual search. In consequence, non-foveated small items may appear too small in the periphery to be learned visually as context.



**Figure 3.** (a) Stimuli used by Chun & Jiang (1998), Olson & Chun (2002), Jiang & Wagner (2004), Lleras & von Mühlenen (2004), and Brady & Chun (2007). Observers searched for ‘T’ among ‘L’s. (b) Stimuli used by Chun & Jiang (1999). Observers searched for a shape symmetric around the vertical axis. (c) Stimuli used by Brockmole, Castelhana, & Henderson (2006). Observers searched for ‘T’ or ‘L’ that was always located on the room table. (d) Stimuli used by Jiang et al. (2006). Observers searched for ‘T’ among ‘L’s. (e) Stimuli used by Torralba et al. (2006). Observers searched for pedestrians. (Figures reprinted with permission from each study).

Spatial context is not necessarily represented as a global gist pattern in memory. It can also take effect via an ensemble of pairwise positional associations between a target and

accompanied distractors. Indeed, spatial cueing effects can be obtained from a novel spatial configuration by combining individual locations predictive of target position (Jiang & Wagner, 2004, Experiment 1). In addition, a positive correlation between set size and cueing effects was observed (Chun & Jiang, 1998, Experiment 4), indicating that spatial cueing can integrate across pairwise associations. Thus, the strength of cueing effects may depend on correlations of individual target-distractor pairs, which sometimes differ from the correlation between a target and its co-varied context as a whole. In fact, training with crowded search displays can diminish cueing effects (Hodsoll and Humphreys, 2005) because a decreasing target-distractor correlation may occur with an increasing probability of a distractor location being recycled in various contexts where it is paired with different target locations. Similarly, small cueing effects that result when a context primes more than one target location (Chun & Jiang, 1998, Experiment 3) can be intensified by more training to strengthen correlations between fixed target-distractor locations (Chun & Jiang, 1998, Experiment 6). Finally, a target-distractor relationship is not learned until a target is found at the end of a search trial, and in this case target presence is necessary to gain spatial cueing effects (Kunar & Wolfe, 2009).

It is worth noting that Ogawa & Watanabe (1997) replaced a searched layout by an unsearched context right before a target fixation and reported that both contexts facilitated later searches. This study, however, does not imply that spatial context is learned before a target is fixated, because the short-term memory (STM) trace of a searched layout after abrupt context substitution may persist for a short period during which target-triggered learning encodes both available context layouts into long-term memory (LTM).

Aside from statistical regularities in external search displays, internal factors such as attention also regulate memory encoding and retrieval of spatial context. Jiang & Chun (2001) showed that invariant configurations of an attended color evoked stronger spatial cueing effects than the ones of an unattended color in the same search display. Lleras & von Mühlenen (2004) used the spatial cueing paradigm with different instructions to bias searchers' strategies, and found that 80% of subjects in the passive/receptive group showed context-induced search RT decreases down to around 300ms, whereas 65% of subjects in the active/deliberate group showed context-induced search RT increases up to around 100ms. It is curious why negative cueing effects, or increased search RT, arose from an informative spatial context. In any case, attention modulates the efficacies of contributions from equally predictive context locations, and may play a role in making local context more effective than global/distant context for predicting target locations (Olson & Chun, 2002; Brady & Chun, 2007) or vice versa (Brockmole, Castelano, & Henderson, 2006). In summary, target-distractor associations and attentional modulation are two key components in contextual cueing.

ARTSCENE Search clarifies how both types of mechanisms work and interact during evidence accumulation and memory-based contextual guidance in visual search. To be more precise, ARTSCENE Search synthesizes three types of attentional factors, each of which interdependently contributes to the process of spatial selection based on its own inputs: Bottom-up inputs attract attentional saliency directly from a visual scene; gist-based top-down spatial attention learns to prime target positions from correlated locations; and feature-based top-down object attention learns to prime the identity and features of a target from correlated distractors. In the next two sections, neural data will be reviewed to suggest how the brain accommodates these three attentional processes to achieve visual search, context learning, and scene understanding in general.

#### **4. Brain Systems for Scene Understanding and Visual Search**

Visual scenes are processed in two major interactive pathways. The ventral What cortical processing stream carries out object perception, recognition and prediction, whereas the dorsal Where cortical processing stream carries out target selection and action in space (Ungerleider & Mishkin, 1982; Goodale & Milner, 1992). Evidence from context-dependent object recognition suggests that the low spatial frequency components of a scene are rapidly transmitted through magnocellular projections before the high spatial frequency counterparts are available for object recognition in the parvocellular pathway (Bar et al., 2006; Kveraga, Boshyan, & Bar, 2007). The magnocellular stream is thus likely to extract aspects of scene gist and trigger top-down priming for object recognition or visual search, whereas the parvocellular stream is better suited for processing detailed featural information in an object or a scene.

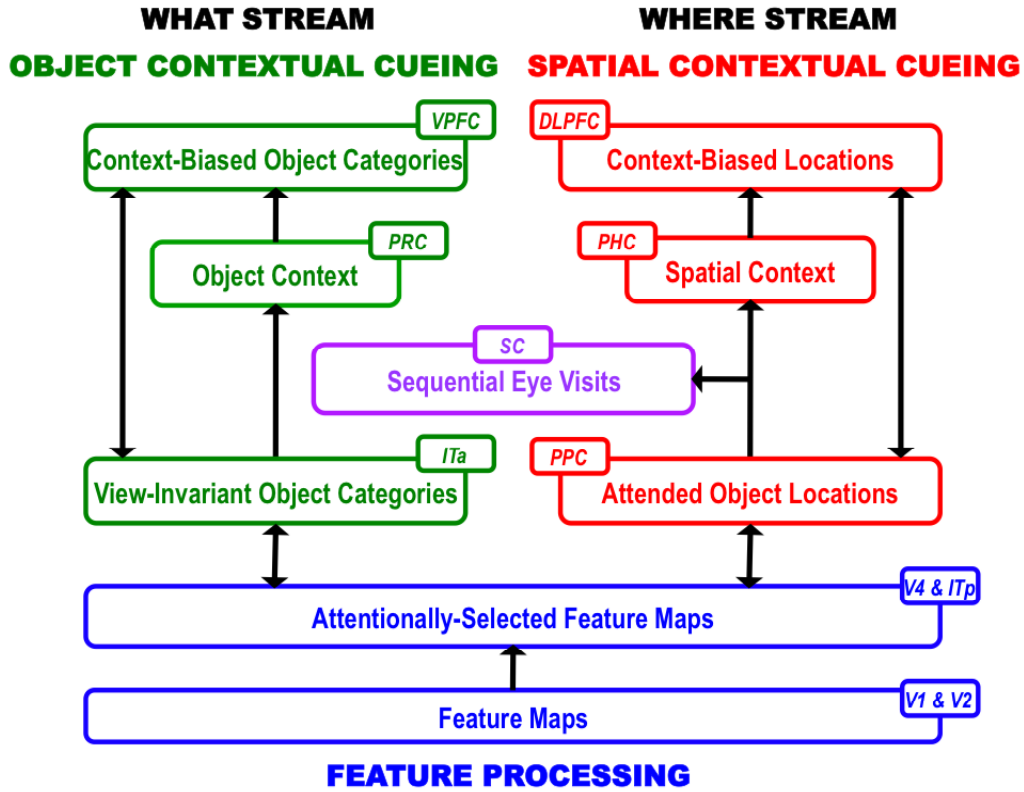
Top-down priming occurs in both cortical streams, and enhances effective contrast of an attended stimulus (Grossberg, 1980, 1999; Carrasco, Penpeci-Talgar, & Eckstein, 2000; Reynolds & Chelazzi, 2004). For the ventral What stream, studies of color-based attention found attentional modulation not only in inferotemporal cortex (ITC) and V4, but also in early visual areas LGN, V1, V2 and V3 (Grossberg & Mingolla, 1985; Sillito, Jones, Gerstein, & West, 1994; Saenz, Buracas, & Boynton, 2002; Müller et al., 2006). ITC, among other areas, due to its direct feedback projections to V4, V2, and V1 (Rockland & Drash, 1996). As for the dorsal Where stream, the posterior parietal cortex (PPC) engages in both spatial shifts of attention and non-spatial tasks such as feature conjunction of shapes and textures (Wojciulik & Kanwisher, 1999), as a part of frontoparietal attention network (Egner et al., 2008). Lateral intraparietal area (LIP), especially, is believed to represent a feature-sensitive saliency map and to guide the selection of saccadic eye movements (Vidyasagar, 1999; Gottlieb, 2007; Buschman & Miller, 2007). Patients with bilateral parietal lesions lose the ability to spatially localize objects, and they perceive illusory conjunctions even with long displays of only two objects (Treisman, 2006).

Prefrontal cortex (PFC) displays persistent activities in the delay period of working memory tasks (Fuster & Alexander, 1971; Fuster, 1973; Funahashi, Chafee, & Goldman-Rakic, 1993; Miller, Erickson, & Desimone, 1996; Chafee & Goldman-Rakic, 1998; Sakai, Rowe, & Passingham, 2002; Curtis & D'Esposito, 2003), and provides top-down priming in many experimental tasks (see reviews by Miller & D'Esposito, 2005). Dorsolateral PFC (DLPFC, area 46 in the convention of Petrides, 2005) influences spatial selection of targets (Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000) and plans sequences for saccades in a particular context (Averbeck & Lee, 2007). In contrast, ventral PFC (VPFC, area 47/12 in the convention of Petrides, 2005) showed stronger fMRI BOLD signals for visual objects that are highly associated with a certain context (e.g., an oven) than objects that are not paired with any unique context (e.g., a hat), and was more active in successful than unsuccessful attempts of object recognition (Bar et al., 2006). Frontal cortices including DLPFC and the frontal eye fields (FEF) may mediate target biasing in contextual cueing through reinforcement learning gated by dopamine (cf., Brown, Bullock, & Grossberg, 2004; Schultz, 2006), consistent with the role of DLPFC in goal-directed behavior (Miller & Cohen, 2001; Fuster, 2008) and FEF in visual target selection (Schall & Thompson, 1999; Buschman & Miller, 2007).

Connecting with PFC, the medial temporal lobe (MTL) plays an essential role in processing context in a scene by which a target can be quickly defined. In the MTL, hippocampus is a major component encompassed by parahippocampal regions, including complementary roles for parahippocampal cortex (PHC) and perirhinal cortex (PRC) in spatial and object contextual processing. Chun & Phelps (1999) reported that amnesic patients with

damage in medial temporal lobe did not exhibit spatial cueing effects seen in the control group. Manns & Squire (2001) further showed that spatial cueing effects could be impaired by extensive damage to the MTL along with variable damage to lateral temporal cortex, but not by damage confined to the hippocampal formation. This may explain why memory consolidation after sleep, believed to be hippocampus-dependent, is not seen in spatial cueing effects (Mednick, Makovski, Cai, & Jiang, 2009). Using fMRI, Aminoff, Gronau, & Bar (2007) examined context learning in a passive viewing paradigm, and demonstrated that the spatial-context condition elicited more activation in the posterior PHC (i.e., the parahippocampal place area, PPA) than the no-context condition, whereas the object-context condition elicited more activation in the anterior PHC and its adjacent PRC than the no-context condition. Also by fMRI, Jiang, King, Shim, & Vickery (2006) observed the involvement of PPA in scene-based spatial cueing (see Figure 3d). These findings are consistent with data showing that PPA responds more vigorously to structured scenes than to single objects (Epstein & Kanwisher, 1998), and is engaged in coding scene layouts (Epstein, Stanley, Harris, & Kanwisher, 1999) and boundaries (Park, Intraub, Yi, Widders, & Chun, 2007), whereas PRC is implicated in coding stimulus-stimulus associations (Murray & Richmond, 2001; Naya, Yoda, & Miyashita, 2003; Naya, Yoshida, Takeda, Fujimichi, Miyashita, 2003) and high-order feature conjunctions (Murray & Bussey, 1999; Bussey & Saksida, 2002).

Taken together, the above-mentioned neural data suggest a division of labor among several brain areas. ARTSCENE Search illustrates how they interact to overcome their complementary deficiencies to attain efficient memory-based visual search.



**Figure 4.** Macrocircuit of the ARTSCENE Search neural model for visual context processing. **V1**=First Visual Area or Primary Visual Cortex; **V2**=Second Visual Area; **V4**=Fourth Visual Area; **PPC**=Posterior Parietal Cortex; **IT<sub>p</sub>**=Posterior Inferotemporal Cortex; **IT<sub>a</sub>**=Anterior Inferotemporal Cortex;

**PHC**=Parahippocampal Cortex; **PRC**=Perirhinal Cortex; **DLPFC**=Dorsolateral Prefrontal Cortex; **VPFC**=Ventral Prefrontal Cortex; **SC**=Superior Colliculus.

## 5. The ARTSCENE Search Model

Figure 4 provides a macrocircuit of the ARTSCENE Search model in terms of key brain areas involved in scene perception and scene memory. In Figure 4, visual areas from V1/V2 to V4/IT<sub>p</sub> compose basic scene percepts; IT<sub>a</sub> integrates both bottom-up and top-down information for invariant object recognition; PPC integrates both bottom-up and top-down information for selection of target location; SC directs eye scans in a scene; medial temporal lobe areas (PRC and PHC) encode object and spatial context, respectively; and prefrontal cortex (areas VPFC and DLPFC) mediate context-biased target-based or location-based representation in working memory.

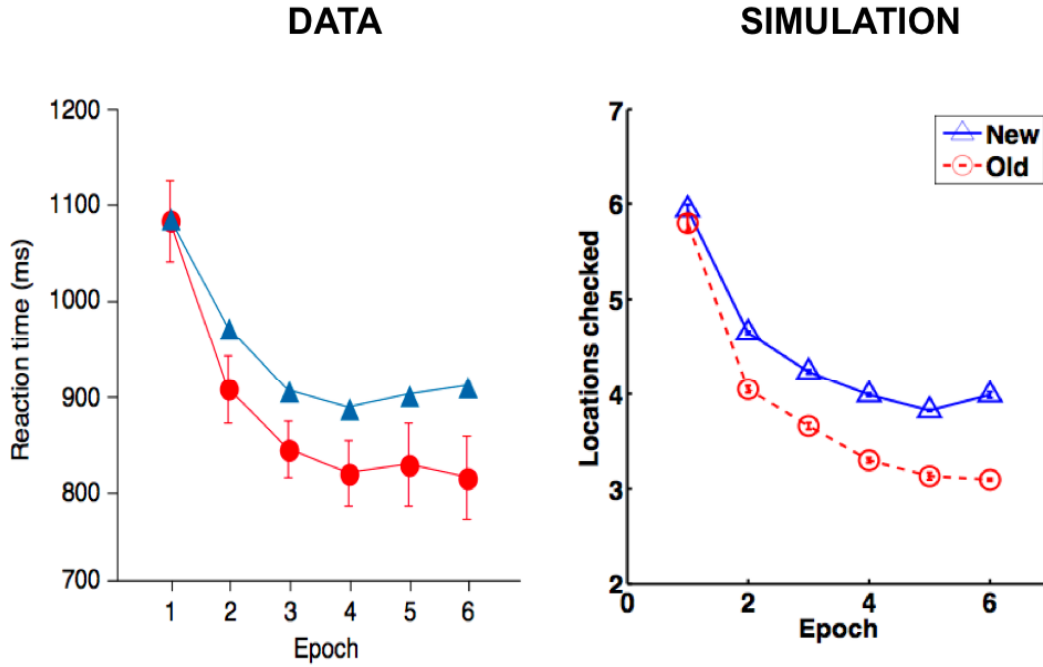
The model simplifies the descriptions of visual preprocessing, in particular, to enable quantitative simulations of system interactions at processing levels which are rate-limiting in controlling contextually guided visual search. That is why the processes V1/V2 and V4/IT<sub>p</sub> are lumped together and only the minimum computations simulated that are needed to input to the higher processing levels whose dynamics are sufficient to quantitatively simulate the targeted data.

Structurally, ARTSCENE Search integrates computations within and between the cortical What and Where streams. As illustrated in Figure 4, the early What-Where segregation is further extended into MTL (see reviews by Eichenbaum, Yonelinas, & Ranganath, 2007) and PFC (Levy & Goldman-Rakic, 2000). In MTL, PHC reciprocally projects to PPC, and PRC reciprocally projects to IT<sub>a</sub> (Suzuki & Amaral, 1994). In PFC, DLPFC (area 46) links to PPC directly and to PHC via the retrosplenial cortex (RSC), whereas VPFC (area 47/12) strongly links to IT<sub>a</sub> (area TE) and PRC (Petrides, 2005). The anatomical segregation of two visual streams reflects on the physiological differences of brain areas involved in context processing. Functionally, the What and Where streams process object and spatial aspects of a scene, respectively (Grossberg, 2000a). ARTSCENE Search proposes that the Where stream computes scene layout for spatial cueing in early gist-based scene analysis, and the What stream computes features in a scene for object cueing. In particular, DLPFC primes possible target locations in PPC based on attention-modulated spatial context, or spatial scene gist, represented in PHC, whereas VPFC primes possible target categories in IT<sub>a</sub> based on the object context, or history of viewed objects, represented in PRC. This model treatment of PRC is consistent with the fact that rhinal (including entorhinal and perirhinal) cortex is necessary for monkeys to learn the temporal relationship between a visual cue and a deferred reward (Liu, Murray, & Richmond 2000), which requires STM of viewed cues to be maintained until reward delivery to establish the cue-reward associations.

The mathematical equations and parameters of the model are defined in the Appendix. Simulations of ARTSCENE Search use the experimental parameters described in each study, including the matrix size of a search display, and the number of search objects, trials, blocks and epochs. These parameters are summarized in Table 1. Since the model does not factor in all possible biological processes that contribute to search time costs, such as stimulus registration and the dynamics of saccade execution, the search reaction time (RT) in a trial,  $t$ , is instead simulated by the number of locations checked to discover a target in a discrete display matrix. That is, an RT reduction in search is attributed to a search in which a smaller number of saccades occurs, as in the experiments of Tseng & Li (2004) and the model of Brady & Chun (2007).

## 6. Simulation Results

ARTSCENE Search provides a unified explanation about data about spatial/object, distant/local, and positive/negative contextual cueing effects. Specifically, spatial cueing and object cueing effects illustrate learned correlations between a target and its context objects in location or appearance (i.e., the interaction kernels  $W_{xyij}^{HD}$  in Equation 15 or  $W_{nm}^{RV}$  in Equation 22), respectively. In addition, the effectiveness of global versus local cues can be attributed to associative learning resulting from attentional selection of specific context cues (i.e., the spatial layout  $H_{ij}$  in Equation 15). A peripheral object, if salient enough due to its brightness, color, or size from the neighboring background, can draw attentional resources away from other context objects in the parafovea. As a result, a peripheral object can be more viable than a parafoveal object for saliency-dependent associative learning (Equation 15), and vice versa. Both positive and negative spatial cueing can also arise due to different extents of spatial attention centering a search target. A more widely spread attentional window (i.e., a larger  $\sigma_A$  in Equation 14) allows more context objects to be associated with the target, albeit possibly with less attention to each object, which may lead to stronger positive cueing effects. In contrast, a strictly form-fitting attentional window (i.e., a small  $\sigma_A$ ) or shroud (Fazl, Grossberg, & Mingolla, 2009) on a single target cannot establish associations with any context objects, so that no significant contextual effects should be observed. In this case, negative cueing effects can occur at the single subject level whereby all non-target objects become distractors during search.

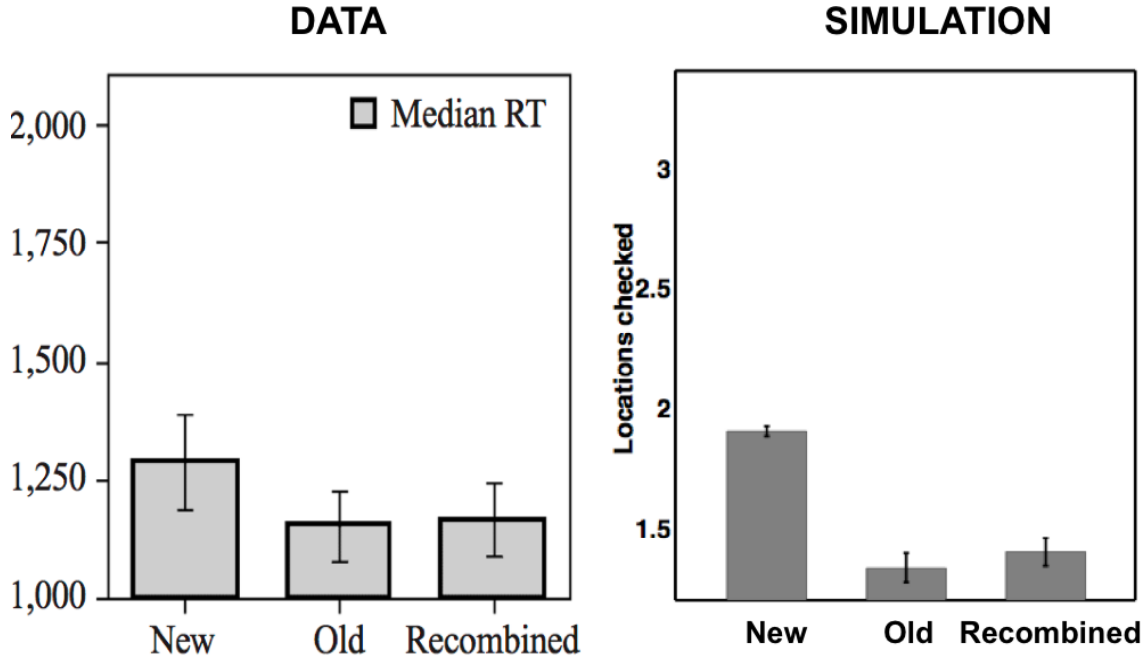


**Figure 5.** Positive spatial cueing effects are the RT reductions for search in a familiar spatial context (i.e., red curves) compared to a new context (i.e., blue curves). The x-axis represents *training time* and the y-axis represents *search time* in a trial. (Data reprinted with permission from Chun, 2000).

For the positive spatial cueing effects (Figure 5), the RT reductions for novel displays reflect the learning of limited target locations in space, and the further RT reductions for repeated displays come from the facilitation of a familiar context layout. ARTSCENE Search replicates



such findings of spatial cueing effects through learning of pairwise associations between locations in the spatial context and target objects. Specifically, when a search display is presented, the layout of search items forms a spatial scene gist, which activates PPC (Equation 9) and its downstream PHC (Equation 12). Each location representation in PHC ( $H_{ij}$  in Equation 12) then learns to vote for its correlated target locations in DLPFC through learned connections from spatial context ( $W_{xyij}^{HD}$  in Equation 15), collectively building up a spatial representation in DLPFC about the likelihood of seeing a target at each location ( $D_{ij}$  in Equation 13). After that, the bottom-up saliency in PPC ( $P_{ij}$  in Equation 9) is modulated by feedback from DLPFC to enhance possible target locations given the current scene layout. As a consequence, strong top-down guidance from DLPFC can alter an eye-scan path that is originally driven only by bottom-up saliency. Importantly, the strongest pairwise association is typically from a target location to itself due to perfect self-correlation, which causes the search RT to drop in the course of training using old target locations in combination with new spatial contexts (e.g., the blue curve in Figure 5).



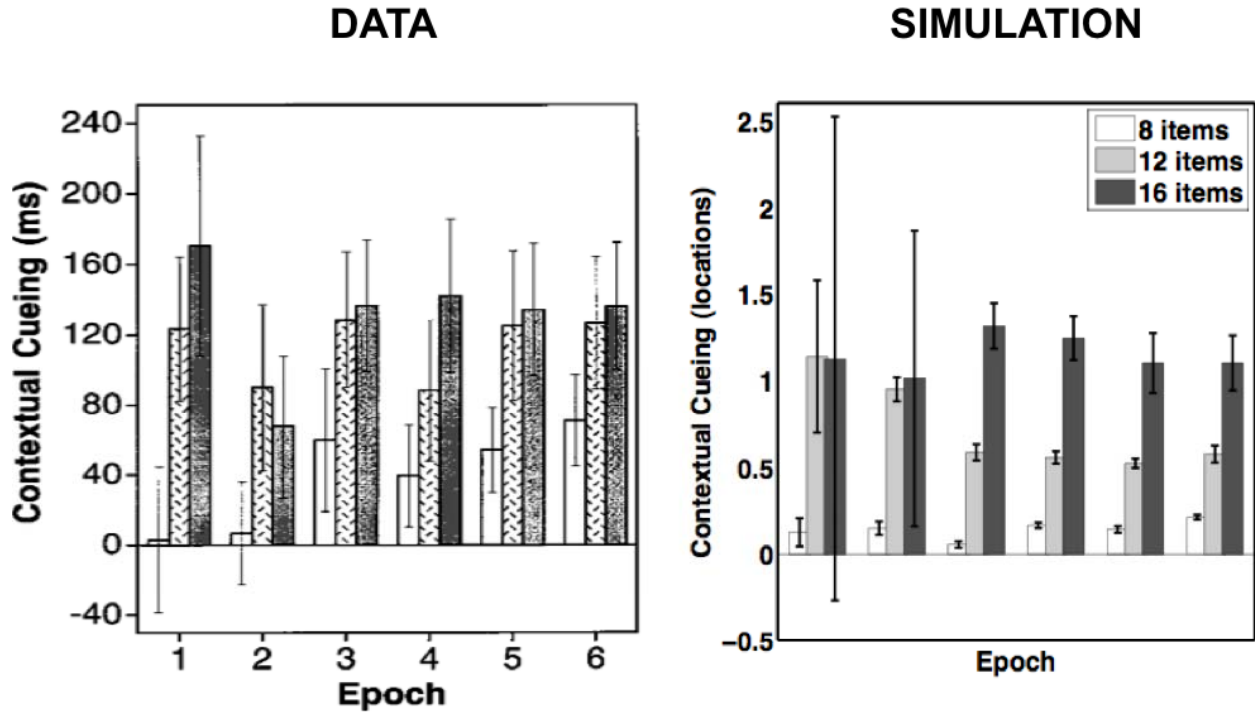
**Figure 6.** Spatial contextual cueing can be obtained from a novel context configuration consisting of predictive individual locations. In the graphs, the y-axis represents *search time* in a trial, and the ‘New’, ‘Old’, and ‘Recombined’ conditions in order refer to search in a novel, familiar or recombined configuration. The recombined configuration is a half-half blend of two ‘Old’ configurations that are equally paired with the same target location during training. (Data reprinted with permission from Jiang & Wagner, 2004, Experiment 1).

In ARTSCENE Search, each occupied location in the search display is a piece of evidence for correlated target locations, and spatial cueing is simply a process of aggregating such location evidence for context memory retrieval (see term  $\sum_{xy} [H_{xy} \phi_{xyij}(\sigma_A)] W_{xyij}^{HD}$  in Equation

13). In consequence, a novel spatial scene gist in model PHC can still drive certain location representations in DLPFC to generate spatial cueing effects as long as individual locations in the

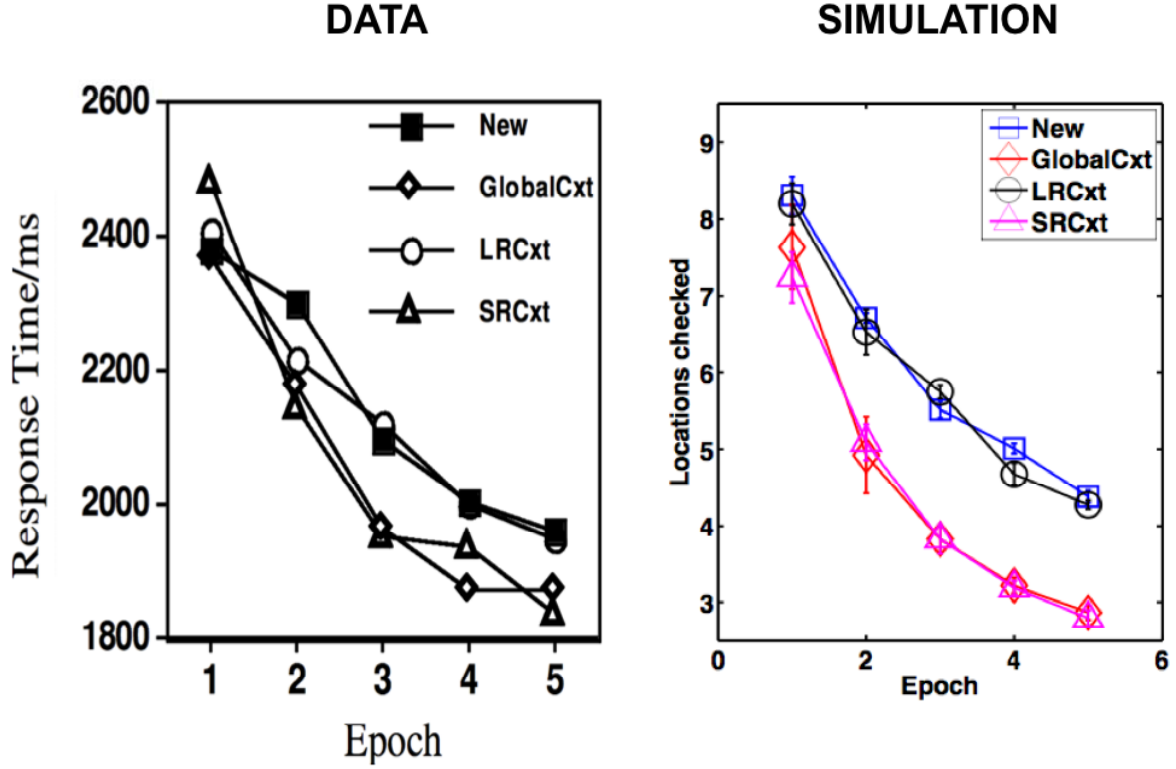


layout have consistently learned to predict a target location (Figure 6). Furthermore, when more congruous evidence is available, the spatial cueing effect is stronger (Figure 7).



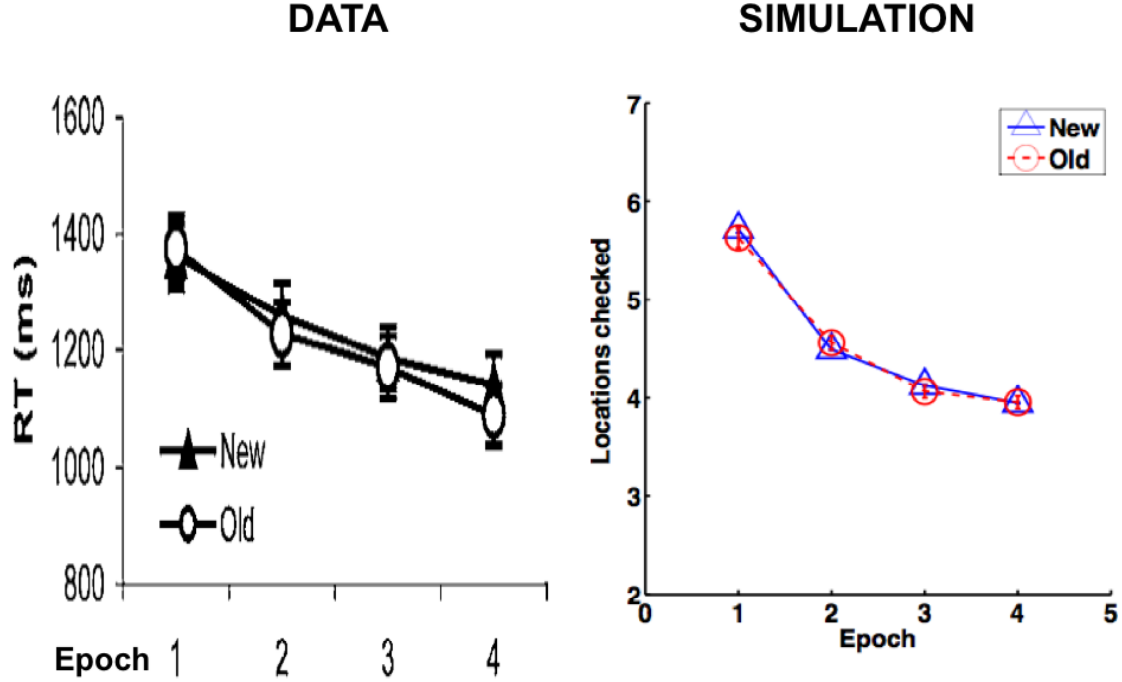
**Figure 7.** Set size effects in spatial contextual cueing. The context-induced RT reductions on the y-axis are more pronounced when the search set size is larger. The x-axis represents *training time*. (Data reprinted with permission from Chun & Jiang, 1998, Experiment 4).

Locations equally predictive of a target location disproportionately contribute to spatial cueing effects, which cannot be solely explained by associative learning. Olson & Chun (2002) compared spatial cueing effects from contexts in separate visual hemifields (left vs. right or upper vs. lower), and found that an invariant short-range context in the target hemifield retained strong cueing effects, as if the whole background context was maintained, but an invariant long-range context in the opposite hemifield yielded no cueing effects, as if the whole background context was a novel configuration (Figure 8; also see Kingstone, Enns, Mangun, & Gazzaniga, 1995 and Alvarez & Cavanagh, 2005 for hemifield differences in visual search and attentional tracking). A follow-up study by Brady & Chun (2007) strengthens this finding by showing an invariant context in the target quadrant alone is as effective as a fully repeated layout. Such results may derive from the allocation of spatial attention, which down-regulates the effectiveness of contexts in the periphery when observers discover and fixate at the target.



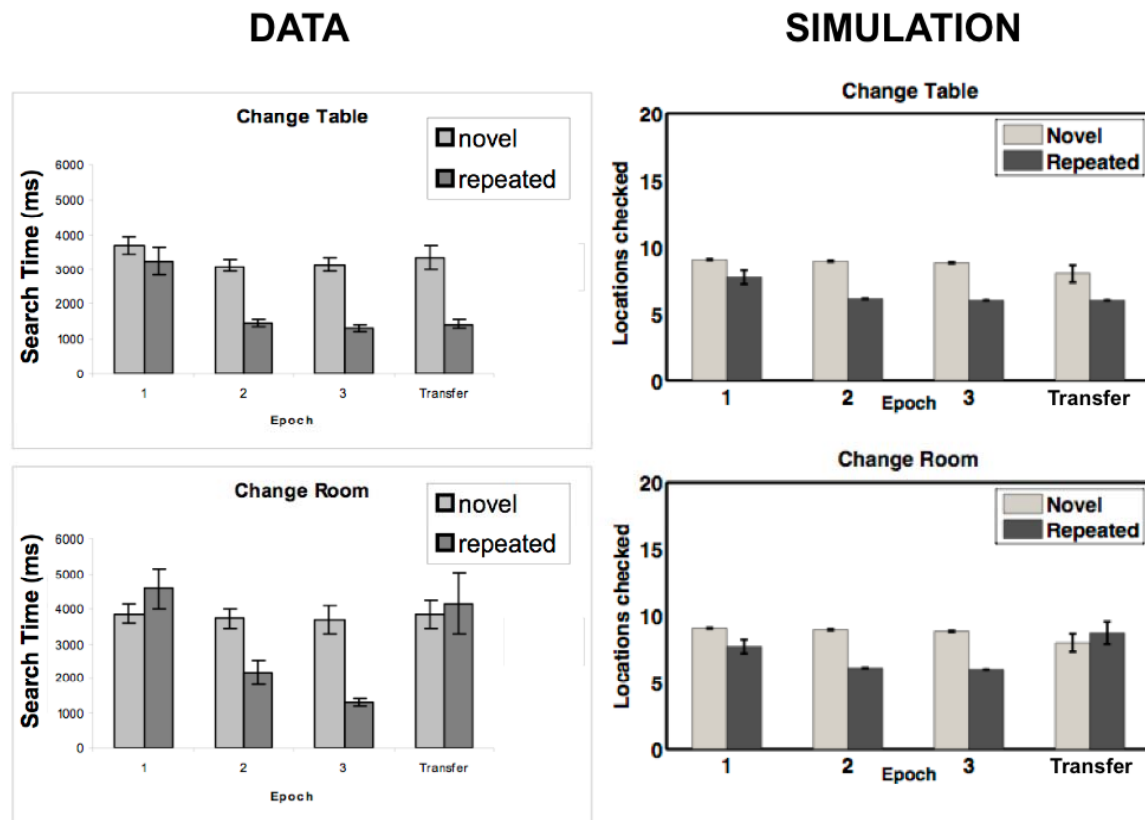
**Figure 8.** Spatial cueing effects can be mainly attributed to predictive locations closer to the target, such as those in the same visual hemifield. In the graphs, the conditions ‘New’, ‘GlobalCxt’, ‘LRCxt’, and ‘SRCxt’ refer to novel, repeated, long-range, and short-range spatial contexts, respectively, with respect to the target location. The x-axis represents *training time* and the y-axis represents *search time* in a trial. (Data reprinted with permission from Olson & Chun, 2000, Experiment 2).

In ARTSCENE Search, the extent of spatial cueing is attained by a Gaussian window ( $\phi_{xyij}(\sigma_A)$  in Equation 14) which limits the spatial context in PHC that modulates DLPFC during context learning (Equation 15) and retrieval (Equation 13). When spatial attention is sharply focused, the Gaussian window only includes the target location and eliminates learning of all other context locations. In this case, the group average of search RT is comparable in both novel and familiar scenes (Figure 9), and negative cueing effects can arise at the single subject level as statistical fluctuations toward RT increase with respect to the group average (i.e., null effect of contextual facilitation).



**Figure 9.** Negative cueing effects or context-induced search RT increases can arise at the single subject level due to focused attention. At the group level in which search RTs are averaged across subjects, there is no significant RT difference for search in a familiar spatial context (red curve with circles) or a novel one (blue curve with triangles). The x-axis represents *training time* and the y-axis represents *search time* in a trial. (Data reprinted with permission from Lleras & von Mühlenen, 2004, Experiment 3).

In ARTSCENE Search, attentionally salient locations constitute effective spatial context. This model property reconciles seemingly conflicting data on what has been learned as spatial context. Opposite to the locality observations in discrete letter displays (Olson & Chun, 2002; Brady & Chun, 2007), Brockmole, Castelano, & Henderson (2006) reported that global or distant context leads to stronger cueing effects than local context in naturalistic scenes. Natural or naturalistic scenes are much richer inputs to the vision system than simple search displays, and consist of high-order textures interfacing adjacent objects (Grossberg & Huang, 2009). Nonetheless, it is unlikely that neural mechanisms for visual search change in response to varying visual stimuli. But if the same search mechanisms are assumed, then the data of Brockmole et al. (2006) are perplexing and seriously challenge the locality model proposed by Brady and Chun (2007). Note, however, the global or distant context (i.e., room objects such as a sofa) in their study is visually much more compelling than the local context (i.e., table) in terms of size and color (see Figure 3c), which are attributes known to capture attention (Wolfe & Horowitz, 2004). It is possible that distant context still outweighs local context even under the peripheral down-regulation of spatial attention. Using discrete object representations in a grid search matrix (see Appendix), ARTSCENE Search simulated this possibility and replicated both distant and local cueing with the same mechanism (Figures 8 and 10). In the simulations, the strength of associative weights between DLPFC and PHC ( $W_{xyij}^{HD}$  in Equation 15) is proportional to the location saliency represented in PHC ( $H_{xy}$  in Equation 15). Accordingly, salient spatial context, no matter where it is in a scene, always strongly predicts the target location.

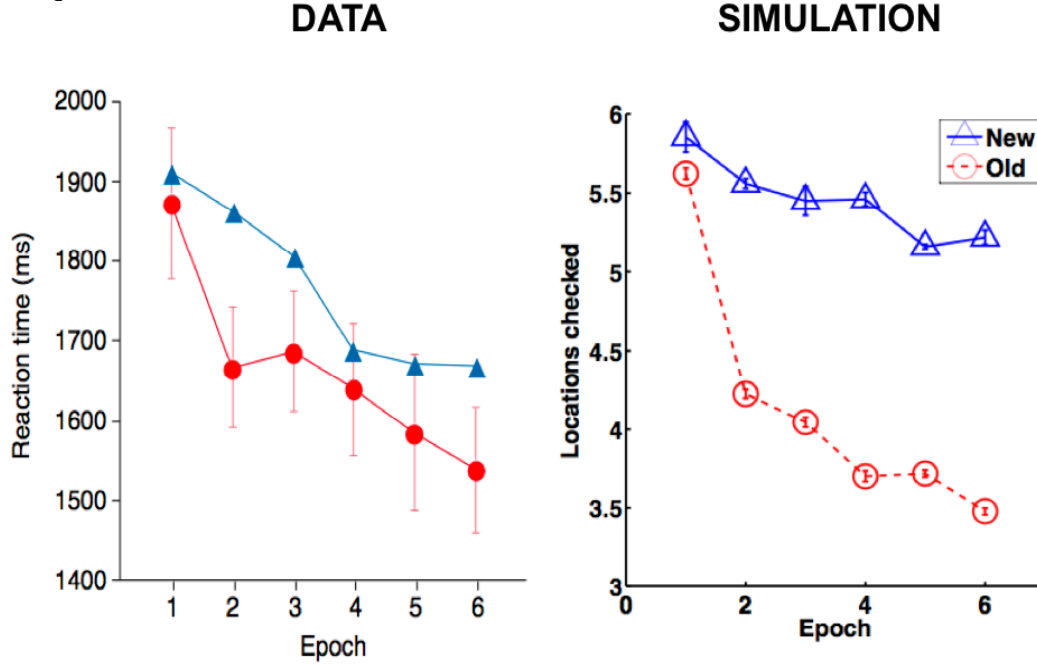


**Figure 10.** Spatial cueing effects due to predictive global/distant context. The graphs show RTs in the three learning epochs and the transfer block during which subjects searched for a target letter on a table (i.e., local contexts) in a furnished room (i.e., global/distant contexts) but with either the familiar table or room changed (upper and lower panels, respectively). The x-axis represents *training time* and the y-axis represents *search time* in a trial. (Data reproduced with permission from Brockmole, Castelano, & Henderson, 2006, Experiment 1).

Target-distractor associations also play a role in the object domain. Chun & Jiang (1999) constructed a semantic context from objects of novel shapes (see Figure 3b), and found that object context alone, without any spatial regularities, could facilitate visual search. In their object cueing experiment, a target is a shape symmetric around the vertical axis in the search display. In the congruent condition, the target is consistently paired with a specific set of distractors across blocks. In the incongruent condition, the target is randomly paired with various sets of distractors. The RT curves of both conditions parallel the ones seen in the spatial cueing effects (Figure 11). In particular, RT reductions for incongruent context reflect learning of specific target features, and the further RT reductions for congruent context come from the learned associative links from distractors to target identities.

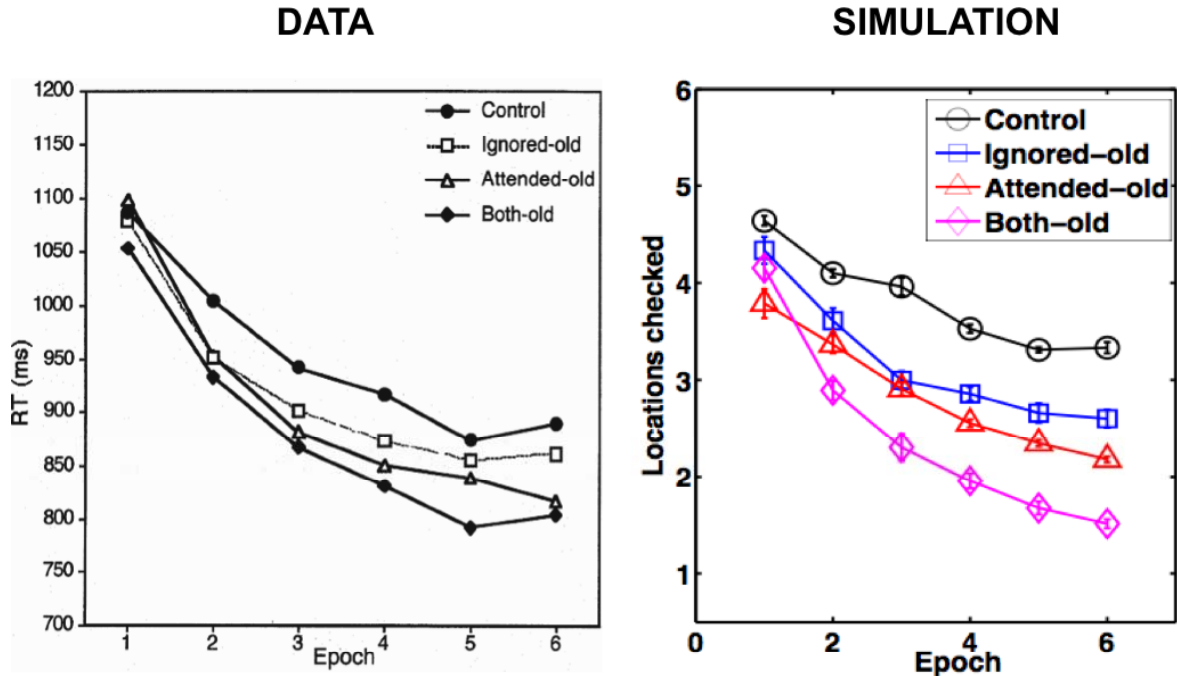
In ARTSCENE Search, object cueing resembles spatial cueing but requires foveal object recognition to establish inter-object rather than inter-location associations. Object cueing is implemented in the ventral What stream and is an inherently sequential process due to a series of eye fixations, each of which binds individual features at the attended location into an integrated object representation (Treisman & Gelade, 1980) for further storage in visual working memory (Luck & Vogel, 1997). In terms of search dynamics, when a search display comes on, the most

salient location is selected for focal attention in PPC (Equation 8). The corresponding object is then recognized in  $IT_a$  (when  $O_m > 0.5$  in Equation 16) and stored temporarily in PRC to initiate formation of object context (Equation 20). If the foveated object is not a target (i.e., the dopamine burst  $G = 0$  in Equation 19), spatial attention at the selected location is then gradually disengaged due to inhibition of return ( $Q_{ij}$  in Equations 9 and 11) and a new selection cycle resumes (Equation 8).



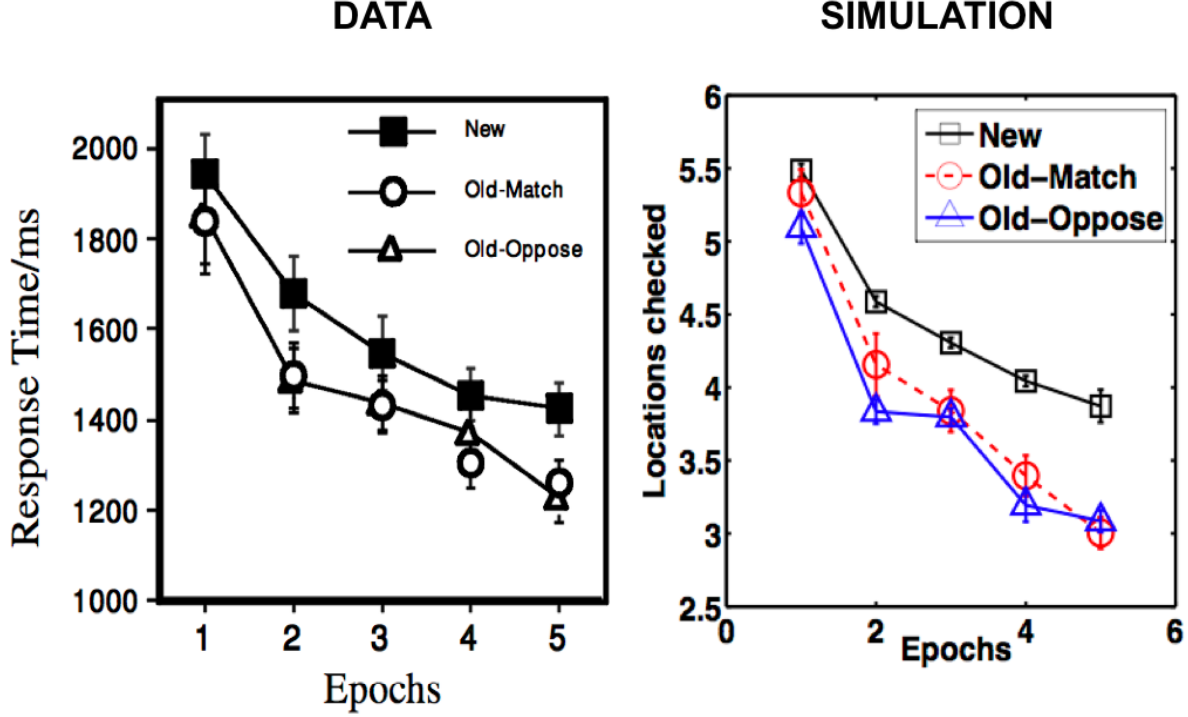
**Figure 11.** Object cueing effects are the RT reductions for search in a congruent object context (i.e., red curves) compared to an incongruent context (i.e., blue curves). The x-axis represents *training time* and the y-axis represents *search time* in a trial. (Data reprinted with permission from Chun, 2000).

During each cycle, a distractor is a new piece of evidence for correlated target identities, and associative votes are carried out through the learned weights from PRC to VPFC ( $W_{nm}^{RV}$  in Equation 22). As the expectation of target identities ( $V_m$  in Equation 21) is built up from the current object context ( $R_n$  in Equation 20), VPFC feeds back to  $IT_a$  ( $V_m$  in Equation 16), which in turn feeds back to  $V4/IT_p$  ( $O_m$  in Equation 6) to enhance target-like features in space. Consistent with data from Vickery, King, & Jiang (2005), the implemented top-down knowledge in visual search primes not only target categories but also visual features of target prototypes. Importantly, while a congruent object context primes a specific target identity (e.g., a butterfly), an incongruent object context non-specifically primes all possible target identities in model VPFC (e.g., objects symmetric around the vertical axis), which still limits the search space and causes RT to drop across trials (see the ‘New’ condition in Figure 11).



**Figure 12.** Selective feature-based attention modulates contextual cueing (Jiang & Chun, 2001, Experiment 3). In the experiment and simulation, a search trial consists of red and green items including the target whose color is maintained and attended throughout. The context layouts are varied in the ‘Control’ condition, but fully repeated in the ‘Both-old’ condition. The ‘Ignored-old’ and ‘Attended-old’ conditions preserve spatial locations across blocks for items of the ignored or attended color, respectively. The x-axis represents *training time* and the y-axis represents *search time* in a trial. (Data reprinted with permission from Jiang & Chun, 2001, Experiment 3).

The ventral What system in ARTSCENE Search not only regulates object cueing but also interacts with the dorsal Where system to modulate spatial cueing. Such What-Where interaction is illustrated when a searcher holds an expectation of target features in mind. One example is from Jiang & Chun (2001), who mixed up the same number of green and red items in a search display and showed that spatial cueing occurs more strongly for the partial contexts that shared the target color than the ones of a different color (Figure 12). In a similar design, Olson & Chun (2002) found that color does not change the effectiveness of a spatial context if color is not a predictive feature for the target (Figure 13). Note that the conditions ‘Control’, ‘Ignored-old’, and ‘Attended-old’ in Figure 12 are the same manipulations in the conditions ‘New’, ‘Old-Oppose’, and ‘Old-Match’ in Figure 13, respectively. However, the ‘Ignored-old’ and ‘Attended-old’ curves separate, whereas the ‘Old-Oppose’ and ‘Old-Match’ curves overlap. This discrepancy originates from the fact that Jiang & Chun (2001) maintained the target color throughout the whole experiment in which participants learned to pay attention only to items sharing the target color, whereas Olson & Chun (2002) had targets colored in red or green half of the trials so that color was an uninformative and ineffective cue for search guidance.



**Figure 13.** Task-irrelevant colors do not affect contextual cueing. In the experiment and simulation, the target color was non-predictable, either red or green. The context layouts are varied in the ‘New’ condition, but preserved across blocks for half items that share the target color in the ‘Old-Match’ condition. In contrast, the ‘Old-Oppose’ condition preserves locations for half items that differ in color from the target. The x-axis represents *training time* and the y-axis represents *search time* in a trial. (Data reprinted with permission from Olson & Chun, 2002, Experiment 4).

ARTSCENE Search replicates both results via feature-based attention. For data from Jiang & Chun (2001), expected target identities in model VPFC prime view-invariant object categories in model  $IT_a$ , which in turn enhance target features such as color represented in model  $IT_p/V4$ . In neural terms, V4 neurons exhibit enhanced responses whenever a preferred stimulus in their receptive field matches a feature of the target (Bichot, Rossi, & Desimone, 2005). In the model, whole-field priming is achieved by a non-specific boost of all item representations within a feature map (Equation 6). Thus, items that share the target color in a scene are processed as candidate targets and also primed. Such What modulation, starting from  $V4/IT_p$  (Equation 6), then propagates along the Where pathway and enhances the corresponding saliency representations of positions in PPC ( $P_{ij}$  in Equation 8) and PHC ( $H_{ij}$  in Equation 12) which, in turn, elevates the efficacy of the attended items for spatial cueing through saliency-dependent associative learning between PHC and DLPFC ( $W_{xyij}^{HD}$  in Equation 15). In the simulation (Figure 12), the VPFC object categories can be activated either by exogenous target inputs or by endogenous object cueing from familiar object contexts. To explain data from Olson & Chun (2002), because VPFC simultaneously primes red and green items, it effectively plays no role in search guidance.

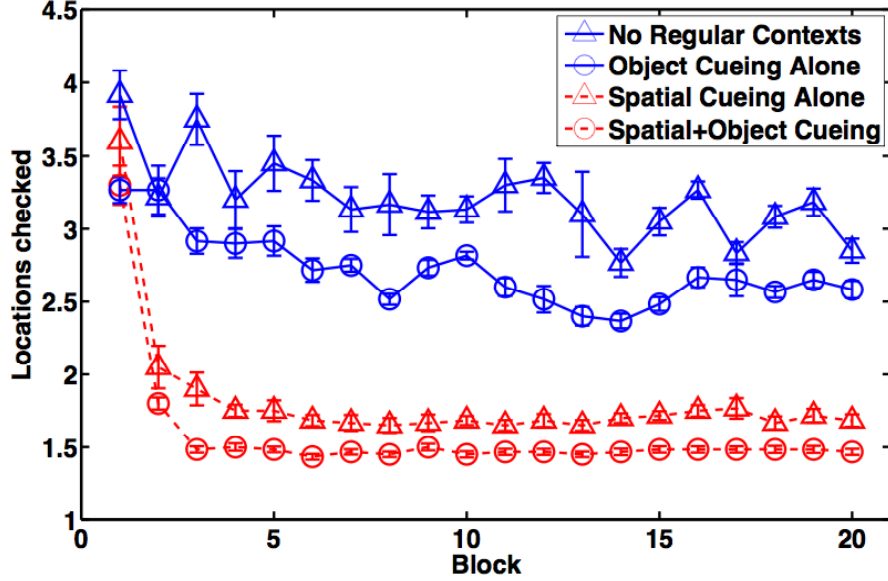
With regard to What-Where interaction, four properties of ARTSCENE Search are worth noting. First, the learned Where-to-Where self-association of a target location can express early in spatial cueing before attentional shifts and eye movements, whereas the learned What-to-What self-association of a target identity cannot express in object cueing because a target is always the



last fixated object in a search trial. Such asymmetry between the target location and identity in contextual cueing may account for reduced search RTs when target locations are fixed and consistently paired with certain distractor identities, but not when target identities are fixed and consistently paired with certain distractor configurations (Endo & Takeda, 2004, Experiment 4). Second, spatial cueing often expresses more strongly than object cueing because spatial cues are collected in parallel due to global gist processing in the early phase of scene analysis and visual search but local object cues are later accumulated in a sequence of eye fixations. Third, given additional sources inputting to model VPFC (e.g., an external input other than the  $IT_a$  and PRC inputs  $O_m$  and  $R_n$  in Equation 21), top-down feature-based attention along the VPFC- $IT_a$ - $IT_p$ /V4 What pathway can modulate bottom-up scene percepts in the V4/ $IT_p$ -PPC-PHC Where pathway to form an effective spatial context in PPC ( $P_{ij}$  in Equation 9) and PHC ( $H_{ij}$  in Equation 12) before eye movements and object cueing occur. Fourth, the degree to which spatial cueing expresses in model DLPFC (i.e.,  $D_{ij}$  in Equation 13) is jointly determined by the saliencies of visually viewed and memory-encoded spatial cues (i.e., PHC activities  $H_{xy}$  and weights of spatial memory  $W_{xyij}^{HD}$  in Equation 13, respectively), both of which are modulated by attention at different times.

For example, a weakly attended and encoded spatial context, coupled with an attention-enhanced retrieval cue, can express strongly. This explains the latent learning phenomenon in which consistent yet ignored spatial cues (i.e., small  $H_{xy}$  and small  $W_{xyij}^{HD}$ ) can barely reduce search RTs during training, but suddenly become effective when attended (i.e., large  $H_{xy}$  and small  $W_{xyij}^{HD}$ ), as seen in the ‘Ignored old’ testing condition of Jiang & Leung (2005). On the other hand, consistent and attended spatial cues, represented in model PHC and stored in the memory weights from PHC to DLPFC, can reduce search RTs during training, but suddenly become less effective when ignored during testing. In this case, the weak RT benefit can be further counteracted by attended yet novel spatial cues that incongruently predict various target locations in model DLPFC, and result in no search facilitation, as seen in the ‘Attended old’ condition of Jiang & Leung (2005).





**Figure 14.** Integrated contextual cueing effects. The simulation used a with-in subject design to show that spatial-plus-object regularities in the search set reduce RT more than the conditions where only spatial or object information is predictive. In the graph, the x-axis represents *training time* and the y-axis represents *search time* in a trial.

Last but not least, in ARTSCENE Search, when both spatial configurations and object identities are predictive of targets, spatial and object cueing can work in concert to provide more accurate guidance than ones acquired from spatial or object cueing alone (Figure 14). This What-and-Where integration benefit is consistent with behavioral data (Gronau, Neta, & Bar, 2008; Endo & Takeda, 2004, Experiment 3) that task RTs are the lowest among conditions when visual stimuli are both spatially and semantically related. In terms of search dynamics, ARTSCENE Search first engages the PPC-PHC-DLPFC circuit to execute gist-based spatial cueing (Equations 9, 12, and 13), and the top-down spatial guidance for targets from DLPFC onto PPC ( $D_{ij}$  in Equation 9) is then integrated with V4/IT<sub>p</sub>-to-PPC bottom-up signals ( $S_{ijk}$  in Equation 9), which are dynamically modulated via the feedback VPFC-IT<sub>a</sub>-IT<sub>p</sub>/V4 circuit by object cueing based on accumulated IT<sub>a</sub>-PRC-VPFC activations after a series of object recognitions. Put differently, in visual search, spatial attention is first distributed to the whole visual field to apprehend the spatial scene gist, which gives rise to a rapid first-order hypothesis of where a target is located. This hypothesis is then incrementally refined by subsequent eye fixations, which identify objects in the scene using focal attention. To summarize, ARTSCENE Search employs the principle of global-to-local visual processing, embodies biologically plausible neural mechanisms, and is capable of guiding attention deployment for most efficient target search based on all the spatial and object regularities in a scene.

## 7. Discussion

ARTSCENE Search reduces to simpler search models given specific conditions. Without any top-down expectations, ARTSCENE Search may be compared to a FIT/SM model (e.g., Treisman & Gelade, 1980; Itti & Koch, 2001), which determines eye-scan paths based purely on bottom-up location saliency. When specific target features are expected before search, ARTSCENE Search functions like Guided Search models (e.g., Wolfe, 1994; Navalpakkam & Itti, 2005). When spatial regularities exist in the environment, ARTSCENE Search generates

spatial priming based on the spatial gist of a scene, which is similar in spirit to models proposed by Torralba et al. (2006) and Brady & Chun (2007). As a synthesis of these important families of visual search models, ARTSCENE Search inherits the explanatory power from these successful models to explain various search phenomena not presented in this article. These include pop-out feature search, serial conjunction search, inefficient feature search, efficient conjunction search, set size effects, basic search asymmetry, as discussed earlier in Section 2.

ARTSCENE Search, however, differs greatly from other search models in several key respects. As opposed to psychological models (e.g., Treisman & Gelade, 1980; Wolfe, 1994; Torralba et al., 2006; Brady & Chun, 2007), ARTSCENE Search is a neural model, which makes specific predictions about what cortical areas and brain mechanisms underlie human search behavior. Beyond functional specification, ARTSCENE Search further simulates the neural dynamics of each model region by differential equations, which enables the model to emulate aspects of brain dynamics in real time. In particular, the model illustrates how various brain regions in the cortical What and Where pathways, including subregions of the model temporal lobe and the prefrontal cortex, dynamically coordinate bottom-up, spatial top-down, and object top-down attention during visual search. Significantly, only with the dynamical interactions among these three attentional systems can ARTSCENE Search concurrently simulate spatial and object cueing effects and reconcile opposite experimental observations of a similar design under the same framework (see Section 6). In addition, due to its unique mechanism of evidence accumulation through eye movements, ARTSCENE Search implements eye fixations as a series of information gathering acts by which the likelihood of seeing a target at every location, or the saccadic plan, can be dynamically revised in the model PPC during the course of search. It thus stands out from other search models that determine a fixed plan of eye movements based on location saliency prior to search, and treat eye fixations as a series of non-target rejections (e.g., Treisman & Gelade, 1980; Wolfe, 1994; Itti & Koch, 2001; Navalpakkam & Itti, 2005; Torralba, Oliva, Castelhana, & Henderson, 2006; Brady & Chun, 2007).

The ARTSCENE Search model clarifies how the brain combines locally uncertain combinations of scenic information, through learning, into predictive decisions for both recognizing scenic type and for more efficiently commanding eye movements to discover a target object within a scene. The model does this without using a Bayesian formalism. Instead, it articulates brain principles and mechanisms that are embodied in hierarchically organized feedforward and feedback interactions within and across the What and Where cortical processing streams, including cortical areas ITC, PRC, and VPFC in the What stream, and PPC, PHC, and DLPRC in the Where stream. In a like manner, Grossberg & Pilly (2008) developed a detailed neural model of how cortical areas V1, MT, MST, LIP, and the basal ganglia interact to make eye movement decisions whose properties quantitatively simulate all the critical data properties of the psychophysical and neurophysiological experiments of Roitman & Shadlen (2002) and Shadlen & Newsome (2001) in response to their probabilistically defined motion stimuli. Thus it seems that a principled understanding of the brain mechanisms that subserve various kinds of perceptual decisions and actions may not be best expressed in terms of Bayesian concepts which do not disclose the underlying brain principles and mechanisms that carry out the decisions. Rather, current brain models articulate new design principles that may be expressed in neural cells, circuits, and systems that may be directly tested on multiple levels of behavioral and brain organization.

Despite its advances, ARTSCENE Search is not yet a complete model of visual search. In terms of representations, the model inputs are discrete search displays where items are point

objects. For the model design to work on real images, future work needs to incorporate mechanisms of boundary and surface processing (Grossberg, 1994) and surface-fitting attentional shrouds (Fazl, Grossberg, & Mingolla, 2009) to attain figure-ground separation and define the boundaries and surfaces of an attended object for further recognition (e.g., Walther & Koch, 2006). Moreover, the treatment of spatial coordinates in ARTSCENE Search is simplified. On the one hand, the locality (Olson & Chun, 2002; Brady & Chun, 2007) and translational invariance (Jiang & Wagner, 2004, Experiment 2) of spatial cueing suggest retinotopic processing along with eye fixations and movements (van Asselen & Castelo-Branco, 2009). On the other hand, a predictive target quadrant leads to no spatial cueing effects if it is moved randomly to a different quadrant across repetitions (Brady & Chun, 2007, Experiment 4), which indicates a spatiotopic coordinate for scene memory, as implemented throughout in ARTSCENE Search. In realistic brain computations, gain fields in PPC can mediate transformations between retinotopic and spatiotopic coordinates (e.g., Fazl, Grossberg, & Mingolla, 2009), and FEF may convert spatial representations of targets in DLPFC from spatiotopic into oculomotor coordinates (Funahashi, Bruce, & Goldman-Rakic, 1989). In addition, future work needs to address why a learned but moderately rescaled spatial configuration can still help reduce search RTs (Jiang & Wagner, 2004, Experiment 2), which may involve multiple-scale filtering (e.g., Cohen & Grossberg, 1987; Grossberg & Huang, 2009) to make the best prediction using all regular information in a scene.

Three possible extensions of ARTSCENE Search are to include more crosstalk between the What and Where pathways, and to more fully model spatiotemporal as well as motivational cueing. Jiang & Song (2005) showed that the expression of spatial cueing can be identity-independent or identity-contingent given different training procedures (e.g., all items are white in training trials vs. all items are either white or black in half of the training trials). Endo & Takeda (2004) suggested that contextual cueing can be attained by associations between target identities and distractor configurations, or between target locations and distractor identities, although their designs confounded these two possibilities with the self-associations of a target location and target identity. More experimental data are needed to clarify these What-Where interactions in contextual cueing to guide the development of future models. Moreover, it has been shown that fixed motion trajectories of distractors further improved search of a moving target whose trajectory is repeated across blocks (Chun & Jiang, 1999, Experiment 2; Ogawa, Watanabe, & Yagi, 2009). Ono, Jiang & Kawahara (2005) reported that predictive spatial context can be carried over in short-term memory to speed up target search during the succeeding trial. Such learning of inter-trial or dynamic regularities involves temporal processing mechanisms beyond the current scope of ARTSCENE Search.

The motivational extension is more straightforward. Note that VPFC (area 47/12) in ARTSCENE Search overlaps in part with the orbitofrontal cortex (see review in Kringelbach, 2005), which is implicated in regulating motivational and emotional processing in conjunction with amygdala (Grossberg, 2000b; Ghashghaei & Barbas, 2002; Schoenbaum, Setlow, Saddoris, & Gallagher, 2003; Dranias, Grossberg, & Bullock, 2008; Grossberg, Bullock, & Dranias, 2008). This additional circuit can drive VPFC using incentive motivational amygdala signals to achieve the voluntary goal-directed visual search that was discussed in Section 1. Also, more elaborated temporal and motivational processing may help explain why spatial cueing effects vanished in a block design where the repeated context condition followed the novel context condition, but occurred with a reverse block order (Jungé, Scholl, & Chun, 2007) or in a conventional intermixed design.

To conclude, ARTSCENE Search presents a biologically predictive neural architecture that unifies bottom-up with top-down attention, spatial with object cueing, and instructed with voluntary search. In the ARTSCENE framework, visual search is a special case of scene understanding that includes mechanisms of global-to-local evidence accumulation, learning, and memory. Finally, ARTSCENE Search can be extended along several different directions to provide a more complete model of object and scene learning, recognition, and prediction, and to thereby advance our understanding of high-level visual cognition of a changing world.

## Appendix

ARTSCENE Search is characterized by the following equations. The activity of each model neuron is defined by a membrane, or shunting, equation (Hodgkin & Huxley, 1952; Grossberg, 1973):

$$\tau \frac{dX(t)}{dt} = -A_X X(t) + [B_X - X(t)]I_{excit}(t) - [C_X + X]I_{inhib}(t). \quad (1)$$

In Equation 1,  $X(t)$  represents membrane voltage;  $\frac{dX(t)}{dt}$  is the rate at which  $X(t)$  changes; parameter  $\tau$  corresponds to membrane capacitance and characterizes cell response time; parameter  $A_X$  corresponds to leakage channel conductance and controls the passive decay rate of  $X(t)$ ; parameters  $B_X$  and  $-C_X$  are reversal potentials bounding  $X(t)$  in the interval  $[-C_X, B_X]$ ; and time-varying conductances  $I_{excit}(t)$  and  $I_{inhib}(t)$  represent, respectively, the total excitatory and inhibitory inputs, which are determined by the model architecture in Figure 4. In the simulations, all differential equations are integrated by the Euler method (Press, Teukolsky, Vetterling, & Flannery, 2007) to dynamically estimate  $X(t)$  at time  $T$ :

$$f = \frac{dX(t)}{dt} \approx \frac{X(T) - X(T-1)}{\Delta T}, \quad (2)$$

where the initial value,  $X(0)$ , is set or reset to zero for each search trial, and the integration time step,  $\Delta T$ , is 0.1 for all model equations.

### Stimuli

Each search trial is specified by an object map  $I_{ij}$  in which zero represents object-absent locations  $(i, j)$ , and a positive integer  $I_{ij} = m$  represents a unique point object at the world-centered location  $(i, j)$ . Since the object cueing experiment by Chun & Jiang (1999) used up to ninety-six novel objects,  $m$  is set from one to one hundred to amply simulate object cueing effects among other experiments. Moreover, targets of a search task are pre-specified before a simulated search session by the object indices  $m$  in the target set  $\Omega$ . In real search experiments, the knowledge of  $\Omega$  or the definition of targets, supplied by task instructions, can either be an object category such as the letter ‘T’, or a set of object categories satisfying an abstract rule such as a shape symmetric around the vertical axis. Pre-specified before simulations, the prototype of each object  $m$  serves as a bottom-up filter matching V4/IT<sub>p</sub> inputs to IT<sub>a</sub> for object recognition, and also a top-down prime from IT<sub>a</sub> back to IT<sub>p</sub>/V4 when model neurons in IT<sub>a</sub> are primed by the corresponding object representations in VPFC. Since  $S_{ijk}$  (Equation 5) and  $O_m$  (Equation 18) represent, respectively, the activities of V4/IT<sub>p</sub> and IT<sub>a</sub> neurons, the interconnection weights between V4/IT<sub>p</sub> and IT<sub>a</sub>, or object prototypes, are denoted by

$$\vec{W}_m^{OS} = (W_{m1}^{OS}, W_{m2}^{OS}, W_{m3}^{OS}, \dots, W_{mk}^{OS}, \dots) = \vec{W}_m^{SO}, \quad (3)$$

where  $k$  indexes feature dimensions. This prototype vector represents the featural composition of an object, obtained from early visual processing. For each  $m$  in most simulations,  $\vec{W}_m^{OS}$  is a 100-dimensional binary vector where 10 components,  $W_{mk}^{OS}$ , are randomly chosen to be 1, and the rest 90 components are set to 0. Effectively, a target in such a set-up carries some unique features while sharing some other features with the distractors. To simulate the cases where color is a major attribute in experimental manipulation (i.e., Figures 12 and 13), random feature assignment is avoided. Instead, only 2 components of  $\vec{W}_m^{OS}$  are set to 1. One component is chosen from the two color dimensions in  $\vec{W}_m^{OS}$  to represent either ‘red’ or ‘green’, and the other is

chosen without replacement from the remaining ninety-eight dimensions to represent a unique shape feature for each object  $m$ .

**V1/V2** ( $f_{ijk}^{(m)}$ )

Starting in the primary visual cortex, boundary and surface properties are computed from visual inputs (Grossberg, 1994). In particular, V1/V2 complex cells are tuned to orientation, among other features, and double-opponent blob cells selectively respond to colors on a surface. Since the model focuses on contextual learning at higher-levels of visual processing, low-level processing in model V1/V2 is simplified into a transformation from an object index  $I_{ij}$  to its 100-dimensional feature representation of object  $m$ :

$$\vec{f}_{ij}^{(m)} = (f_{ij1}^{(m)}, f_{ij2}^{(m)}, f_{ij3}^{(m)}, \dots, f_{ijk}^{(m)}, \dots) \equiv \gamma_m \vec{W}_{I_{ij}}^{OS} = \gamma_m \vec{W}_m^{OS}, \quad (4)$$

where  $\gamma_m$  controls overall saliency of the features located at  $(i, j)$ , and  $k$  refers to a specific value (e.g., vertical) on a specific featural dimension (e.g., orientation). In other words, the V1 cell activity  $f_{ijk}^{(m)}$  is driven by the presence of its preferred feature  $k$  in its receptive field (RF) centered at the world coordinate  $(i, j)$  in response to the object  $m$  at location  $(i, j)$ . In all simulations,  $\gamma_m$  was 1 for all objects with the following exception: To simulate distant cueing effects (Figure 10),  $\gamma_m$  was lowered to 0.1 for targets and distractor objects that surround the target location within the median distance of all possible context locations. This setup is an approximation of the naturalistic stimuli in Figure 3c in that context objects distant from the target are more salient than the ones adjacent to the target.

**V4/IT<sub>p</sub>** ( $S_{ijk}$ )

Model area V4/IT<sub>p</sub> receives bottom-up inputs  $f_{ijk}^{(m)}$  from V1/V2 and top-down primes from anterior inferotemporal cortex (IT<sub>a</sub>). Specifically, the V4/IT<sub>p</sub> cell activity  $S_{ijk}$  is driven bottom-up by the  $k^{\text{th}}$  feature in its receptive field centered at the world position  $(i, j)$  and is modulated top-down by the activities of IT<sub>a</sub> neurons  $O_m$ :

$$\frac{d}{dt} S_{ijk} = -S_{ijk} + (1 - S_{ijk}) s_{ijk} - S_{ijk} \sum_{pq} \Phi_{ijpq}(1) s_{pqk}, \quad (5)$$

where the top-down-modulated excitatory input obeys

$$s_{pqk} = 2 f_{pqk}^{(m)} \left( 1 + \sum_m O_m W_{mk}^{OS} \right). \quad (6)$$

In Equation 6,  $\vec{W}_m^{OS}$  is the template of synaptic connection strengths from the  $m^{\text{th}}$  object primed from IT<sub>a</sub> to V4/IT<sub>p</sub>, and  $\Phi_{ijpq}$  is a 2D Gaussian off-surround kernel characterizing a local neighborhood of iso-feature suppression from adjacent neurons in the inhibitory term of Equation 5:

$$\Phi_{ijxy}(\sigma) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} [(i-x)^2 + (j-y)^2] \right\}. \quad (7)$$

The whole-field featural priming in Equation 6 across all locations  $(i, j)$  from a view-invariant object representation  $O_m$  simplifies the IT<sub>a</sub>-IT<sub>p</sub>-V4 feedback pathway whereby a view-invariant object category in IT<sub>a</sub> primes view- and position-variant object categories in IT<sub>p</sub>, which in turn primes the corresponding features in V4/IT<sub>p</sub> within a specific receptive field (Chang, Cao, & Grossberg, 2009). Computationally, local competitions among model neurons in Equation 5 normalize the output of each feature map into the range of zero to one, and enhance the contrasts of visual inputs in each feature map.

### Where Stream (PPC-PHC-DLPFC):

#### PPC ( $P_{ij}$ )

Before the model triggers an instantaneous saccade to the next object, fixation is maintained at the location  $(I, J)$  where the model PPC cell is most active on the priority map  $P_{ij}$ :

$$(I, J) = \underset{i, j}{\operatorname{argmax}} P_{ij}, \quad (8)$$

where  $(I, J)$  is the location that is chosen by the recurrent competitive dynamics in Equation 9. The model PPC forms a spatiotopic priority map whose activities  $P_{ij}$  pool feedforward inputs  $S_{ijk}$  from V4/IT<sub>p</sub> (Equation 5), and are modulated by top-down attentive feedback projections  $D_{ij}$  from DLPFC corresponding to the world location  $(i, j)$  (Equation 13):

$$\frac{d}{dt} P_{ij} = -.01P_{ij} + (1 - P_{ij}) \left[ .05 \sum_k S_{ijk} (1 + 10D_{ij}) + \psi_{0.3}(P_{ij}) \right] - P_{ij} \left[ \sum_{(x,y) \neq (i,j)} \psi_{0.3}(P_{xy}) + 5\psi_{0.9}(Q_{ij}) \right], \quad (9)$$

where the signal function for thresholded neural responses is defined as

$$\psi_p(x) = \begin{cases} 1, & x \geq p \\ 0, & x < p \end{cases}. \quad (10)$$

In Equation 9, the activities  $P_{ij}$  are contrast-enhanced by a recurrent on-center off-surround network in which  $\psi_{0.3}(P_{ij})$  is the on-center feedback,  $\sum_{(x,y) \neq (i,j)} \psi_{0.3}(P_{xy})$  is the off-surround feedback,

and term  $5\psi_{0.9}(Q_{ij})$  is the inhibition of return on selected locations by negative feedback from STM of visited locations  $Q_{ij}$ , which obeys:

$$\frac{d}{dt} Q_{ij} = (1 - Q_{ij}) \psi_{0.5}(P_{ij}). \quad (11)$$

Computationally,  $Q_{ij}$  is switched on when the corresponding PPC location representation,  $P_{ij}$ , exceeds the 0.5 threshold. It then builds up over time to break the positive feedback loop of the maximum  $P_{ij}$  in Equation 9, initiating a new cycle of location selection. Also, this inhibition of return (IOR) by recurrent negative feedback prevents the network from perseverating on the same choice of locations in later selection cycles (Grossberg, 1978; Ullman & Koch, 1985).

#### PHC ( $H_{ij}$ )

Model PHC spatial category neurons receive one-to-one inputs from the spatial location neurons in model PPC, and store multiple such locations in parallel using a recurrent competitive network with linear feedback signals (Grossberg, 1973):

$$\tau_H \frac{d}{dt} H_{ij} = -.01H_{ij} + (1 - H_{ij})(\lambda P_{ij} + H_{ij}) - H_{ij} \sum_{(x,y) \neq (i,j)} H_{xy}, \quad (12)$$

where  $\tau_H$  is the characteristic response time of  $H_{ij}$ , and  $\lambda$  scales the influence of the excitatory input  $P_{ij}$ . The role of model PHC activity,  $H_{ij}$ , is to preserve the initial order of all location salencies in model PPC (namely  $P_{ij}$ ), which is dynamically changed in the course of sequential spatial selections and inhibition of return (Equations 9 and 11). This property of the PHC STM, in tandem with saliency-dependent LTM encoding of spatial context (Equation 15), allows ARTSCENE Search to explain why attended locations are more effective contexts than non-attended locations (see Figures 8, 10, and 12). To obtain this model property, fast  $\tau_H = 0.1$

ensures that  $H_{ij}$  rapidly converges to equilibrium based on the initial values of  $P_{ij}$ , and small  $\lambda = 10^{-5}$  makes later  $P_{ij}$  inputs only weak perturbations to the stored values in  $H_{ij}$  which can thus maintain a stable representation of spatial scene gist (cf., a small input term generates primacy gradients in Bradski, Carpenter, & Grossberg, 1994).

#### **DLPFC ( $D_{ij}$ )**

The DLPFC activity  $D_{ij}$  at world location  $(i, j)$  is activated by bottom-up input  $P_{ij}$  from PPC as well as by the spatial layout  $H_{xy}$  in PHC, and stored by a recurrent competitive network with linear feedback signals (Grossberg, 1973):

$$\frac{d}{dt}D_{ij} = -D_{ij} + (1 - D_{ij}) \left( 10\psi_{0.5}(P_{ij}) + \sum_{xy} [H_{xy}\phi_{xyij}(\sigma_A)]W_{xyij}^{HD} + D_{ij} \right) - D_{ij} \sum_{(x,y) \neq (i,j)} D_{xy}. \quad (13)$$

In the excitatory term of Equation 13, the winner PPC representation  $\psi_{0.5}(P_{ij})$  is the strongest input due to the multiplicative factor 10, and always drives the corresponding  $D_{ij}$  to be the most active location representation in model DLPFC. A spatial context  $H_{xy}\phi_{xyij}(\sigma_A)$  from PHC is formed by weighting the spatial context input  $H_{xy}$  from PHC with a  $(i, j)$ -centered Gaussian function  $\phi_{xyij}(\sigma_A)$ , which represents an attentional window of size  $\sigma_A$ :

$$\phi_{xyij}(\sigma) = \exp \left\{ -\frac{1}{2\sigma^2} [(i-x)^2 + (j-y)^2] \right\}. \quad (14)$$

Term  $\sum_{xy} [H_{xy}\phi_{xyij}(\sigma_A)]W_{xyij}^{HD}$  is an inner product for matching  $H_{xy}\phi_{xyij}(\sigma_A)$  with  $W_{xyij}^{HD}$ , the long-term memory (LTM) of target-biased spatial context. The learned weights  $W_{xyij}^{HD}$  between location  $(x, y)$  in PHC and location  $(i, j)$  in DLPFC obey the Instar learning rule (Grossberg, 1976), whereby learning is doubly gated by the dominant location representation  $\psi_{0.8}(D_{ij})$  and target-triggered dopamine bursts  $G$  (Equation 19) in model DLPFC (Draniak, Grossberg, & Bullock, 2008):

$$\frac{1}{\mu_D} \frac{d}{dt} W_{xyij}^{HD} = G\psi_{0.8}(D_{ij})(W_{xyij}^{HD} - H_{xy}\phi_{xyij}(\sigma_A)). \quad (15)$$

In Equation 15,  $\mu_D$  is learning rate for spatial contexts in the dorsal Where stream, and  $\sigma_A$  is again the size of the attentional window defining the spatial extent of effective context. Both parameters vary in different simulations (see Table 1). Note that the Instar learning here simultaneously pairs each occupied location in a scene with the target location.



Simulations	Subject Number	Set Size	Search Matrix	Trials x Blocks Training (Transfer)	Spotlight Size $\sigma_A$	Learning Rate $\mu_D / \mu_V$
Fig.5: Positive spatial cueing	15	12	8x6	24x30	0.8	$10^{-3} / 0$
Fig.6: Recombined context	15	11	12x8	36x20 (36x3)	1.2	$10^{-3} / 0$
Fig.7: Set size effects	15	8/12/16	12x8	30	1.2	$10^{-3} / 0$
Fig.8: Local cueing	15	16	12x8	32x20	1.2	$10^{-3} / 0$
Fig.9: Negative spatial cueing	15	12	8x6	24x24	0.1	$10^{-3} / 0$
Fig.10: Global/Distant cueing	15	11	8x6	6x9 (6x1)	5.0	$10^{-3} / 0$
Fig.11: Object cueing	15	11	8x6	16x24	5.0	$0 / 10^{-3}$
Fig.12: Attentional learning	15	16	12x8	24x30	2.0	$10^{-3} / 10^{-3}$
Fig.13: Non-predictive features	15	16	8x12	24x20	2.0	$10^{-3} / 10^{-3}$
Fig.14: What-Where Integration	15	6	8x6	16x20	0.5	$10^{-3} / 10^{-3}$

**Table 1.** Simulation parameters. Three free parameters in the model are the attentional windows size  $\sigma_A$  (Equations 13 and 15), the learning rate for spatial context  $\mu_D$  (Equation 15), and the learning rate for object context  $\mu_V$  (Equation 22). Among the three parameters,  $\sigma_A$  critically determines the efficacy of local, distant, and overall context in spatial cueing. It qualitatively changes the results between experimental conditions (e.g., separate vs. overlapping curves), whereas  $\mu_D$  and  $\mu_V$  only quantitatively change how fast learning curves converge to equilibrium. In some simulations, either  $\mu_D$  or  $\mu_V$  was set to zero to examine how much object or spatial cueing alone can account for the observed contextual learning effects. When both parameters are non-zero, an irregular spatial or object context yields no search facilitation, as if  $\mu_D$  or  $\mu_V$  is zero, respectively.

### What Stream (IT<sub>a</sub>-PRC-VPFC):

#### ITC ( $O_m$ )

The ITC cell activities  $O_m$  are driven by bottom-up object recognition signals  $o_m$  and top-down object primes  $V_m$  from VPFC (Equation 21) in an on-center off-surround competitive network (Grossberg, 1973):

$$\frac{d}{dt}O_m = -O_m + (1 - O_m)(o_m + V_m) - .05O_m \left( \sum_{n \neq m} o_n + \sum_{n \neq m} V_n \right). \quad (16)$$

In Equation 16,  $o_m$  is computed by matching the foveated features  $\sum_{ij} \psi_{0.5}(P_{ij}) \bar{S}_{ij}$  with the  $m^{\text{th}}$  object prototype  $\bar{W}_m^{SO}$  through an inner product:

$$o_m = \bar{W}_m^{SO} \cdot \left( \sum_{ij} \psi_{0.5}(P_{ij}) \bar{S}_{ij} \right) = \sum_k W_{mk}^{SO} \left( \sum_{ij} \psi_{0.5}(P_{ij}) S_{ijk} \right), \quad (17)$$

where  $W_{mk}^{SO}$  is the V4/IT<sub>p</sub>-to-IT<sub>a</sub> object template, and  $\psi_{0.5}(P_{ij})$  selects the V4/IT<sub>p</sub> input  $\tilde{S}_{ij}$  (Equation 5) from the most salient location in model PPC (Equation 9). Thus, the identity of the most active  $O_m$  corresponds to the object  $M$  situated at the currently attended location, where:

$$M = \arg \max_m O_m. \quad (18)$$

When the winner object category,  $M$ , is a target, the dopamine bursts  $G$  are then triggered in PFC to initiate learning (Equations 15 and 22):

$$G = \begin{cases} 1, & M \in \Omega \\ 0, & M \notin \Omega \end{cases}, \quad (19)$$

where  $\Omega$  is the pre-specified target set defined earlier in the Stimuli section. If a target is found (i.e.,  $O_M \geq 0.5$  and  $G=1$ ), the search task is completed and a search trial ends after attention is disengaged from the foveated target (i.e.,  $O_M < 0.5$ ) due to the spatial inhibition of return (Equations 9 and 11).

Note that equations for top-down attention need to provide modulatory excitatory priming as well as off-surround inhibition when they act alone. When they act together with a bottom-up input, they amplify the top-down matched part of the input pattern and inhibit mismatched features. Thus, the on-center part of top-down attention functionally embodies a multiplicative action with bottom-up input. In Equation 16, this is achieved through a proper balance of additive on-center and off-surround inputs which, together with matched bottom-up input, causes the desired gain amplification (cf., Carpenter & Grossberg, 1987, 1991; Grossberg & Raizada, 2000). This mechanism allows attention to enhance target candidates which are not necessarily foveated. In contrast, the explicit multiplication of bottom-up by top-down inputs (cf., Bhatt, Carpenter, & Grossberg, 2007) in Equation 9 avoids spatial selection of no-object locations.

#### PRC ( $R_m$ )

Model PRC object category neurons receive one-to-one inputs from the object category neurons in model IT<sub>a</sub>, and store the STM traces of the sequentially activated object representations  $O_m$  in IT<sub>a</sub> through a linear recurrent competitive network:

$$\frac{d}{dt} R_m = -.01R_m + (1 - R_m)(\psi_{0.5}(O_m) + R_m) - .01R_m \sum_{n \neq m} R_n, \quad (20)$$

where  $\psi_{0.5}(O_m)$  in the excitatory term sets a 0.5 threshold so that only objects at the selected locations (i.e.,  $O_m \geq 0.5$ ), rather than all objects in a scene, form an object context in model PRC. PRC cell activities  $R_m$  exhibit recency effects (Bradski, Carpenter, & Grossberg, 1994) over time whereby recently viewed object cues can be associated with the reward-like target more strongly than earlier ones in ARTSCENE Search. Although the exact form of object STM in PRC is not critical to simulate the behavioral data presented in Section 6, a recency gradient of visual cues in model PRC was simulated in keeping with the observation that cue-triggered motivation for rewards, reflected by task error rates and mediated by rhinal cortex, is progressively stronger toward reward delivery (Liu, Murray, & Richmond, 2000).

#### VPFC ( $V_m$ )

The VPFC activity  $V_m$  for the  $m^{th}$  object is driven by bottom-up inputs  $O_m$  from IT<sub>a</sub> as well as the object context  $R_n$  formed in PRC, and stored by a recurrent competitive network with linear feedback signals:

$$\frac{d}{dt}V_m = -V_m + (1 - V_m) \left( 10\psi_{0.5}(O_m) + \sum_n R_n W_{nm}^{RV} + V_m \right) - V_m \sum_{n \neq m} V_n. \quad (21)$$

In the excitatory term of Equation 21, the winner  $IT_a$  representation  $\psi_{0.5}(O_m)$  is the strongest input due to the multiplicative factor 10, and always drives the corresponding VPFC activity  $V_m$  to be the most active object category in model VPFC; term  $\sum_n R_n W_{nm}^{RV}$  is an inner product for matching the viewed object set  $R_n$  with  $W_{nm}^{RV}$ , the LTM of the target-biased object context; and the learned weights  $W_{nm}^{RV}$  between the  $n^{\text{th}}$  object category in PRC and the  $m^{\text{th}}$  object category in VPFC also obey the Instar learning rule that is doubly gated by the dominant object category  $\psi_{0.8}(V_m)$  and target-triggered dopamine bursts  $G$  (Equation 19) in model VPFC:

$$\frac{1}{\mu_v} \frac{d}{dt} W_{nm}^{RV} = G \psi_{0.8}(V_m) (W_{nm}^{RV} - R_n) \quad (22)$$

where  $\mu_v$  is learning rate for object contexts in the ventral What stream and vary in different simulations (see Table 1). Note that the Instar learning here simultaneously updates pairs of target-distractor associations.

## References

- Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, 16(8), 637-643.
- Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and non-spatial associations. *Cerebral Cortex*, 27, 1493-1503.
- Armory, J. L., & Dolan, R. J. (2002). Modulation of spatial attention by fear-conditioned stimuli: An event-related fMRI study. *Neuropsychologia*, 40, 817-826.
- Averbeck, B. B., & Lee, D. (2007). Prefrontal neural correlates of memory for sequences. *Journal of Neuroscience*, 27(9), 2204-2211.
- Backhaus, A., Heinke, D., & Humphreys, G. W. (2005). Contextual learning in the selective attention for identification model (CL-SAIM): Modeling contextual cueing in visual search tasks. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, 3, 87-87.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., Hamalainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Science*, 103(2), 449-454.
- Bhatt, R., Carpenter, G.A., & Grossberg, S. (2007). Texture segregation by visual cortex: perceptual grouping, attention, and learning. *Vision Research*, 47(25), 3173-3211.
- Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, 308, 529-534.
- Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cueing. *Journal of Experimental Psychology: Human Perception & Performance*, 33(4), 798-815.
- Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, 13(1), 99-108.
- Brockmole, J. R., Castelano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 699-706.
- Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, 17, 471-510.
- Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315, 1860-1862.
- Bussey, T. J., & Saksida, L. M. (2002). The organization of visual object representations: a connectionist model of effects of lesions in perirhinal cortex. *European Journal of Neuroscience*, 15, 355-364.
- 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. (1991). *Pattern recognition by self-organizing neural networks*. Cambridge: The MIT Press.
- Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Research*, 40, 1203-1215.

- Chafee, M. V., & Goldman-Rakic, P. S. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *Journal of Neurophysiology*, 79, 2919-2940.
- Chang, H.-C., Cao, Y., & Grossberg, S. (2009). Where's Waldo? How the brain learns to categorize and discover desired objects in a cluttered scene [Abstract]. *Journal of Vision*, in press. [Vision Sciences Society, Naples, Florida, May 8-13, 2009]
- Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, 46, 4118-4133.
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4, 170-178.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28-71.
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10, 360-365.
- Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic patients with hippocampal damage. *Nature Neuroscience*, 2, 844-847.
- Cohen, M.A., & Grossberg, S. (1987). Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics*, 26, 1866-1891.
- Curtis, C. E., & D'Eposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7, 415-423.
- Dranias, M., Grossberg, S., & Bullock, D. (2008). Dopaminergic and non-dopaminergic value systems in conditioning and outcome-specific revaluation. *Brain Research*, 1238, 239-287.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433-458.
- Egner, T., Monti, J., Trittschuh E., Wieneke C., Hirsch J., & Mesulam M. (2008). Neural integration of top-down spatial and feature-based information in visual search. *Journal of Neuroscience*, 28(24), 6141-6151.
- Eichenbaum, H., Yonelinas, A. R. & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Reviews of Neuroscience*, 30, 123-152.
- Endo, N., & Takeda, Y. (2004). Selective learning of spatial configuration and object identity in visual search. *Perception & Psychophysics*, 66(2), 293-302.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392, 598-601.
- Epstein, R., Stanley, D., Harris, A., & Kanwisher, N. (1999). The parahippocampal place area: Perception, encoding, or memory retrieval? *Neuron*, 23, 115-125.
- 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, 1-48.
- Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61, 331-349.
- Funahashi, S., Chafee, M. V., & Goldman-Rakic, P. S. (1993). Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature*, 365, 753-756.
- Fuster, J. M. (1973). Unit activity in the prefrontal cortex during delayed response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology*, 36, 61-78.

- Fuster, J. M. (2008). *The prefrontal cortex, 4th edition*. Boston: Academic Press.
- Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, 173, 652-654.
- Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience*, 115(4), 1261-1279.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 20-25.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535-574.
- Gottlieb, J. (2007). From thought to action: The parietal cortex as a bridge between perception, action, and cognition. *Neuron*, 53(1), 9-16.
- Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, 20(3), 371-388.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 213-257.
- Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23, 187-202.
- Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.), *Progress in theoretical biology*, Volume 5. New York: Academic Press, pp. 233-374.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1-51.
- Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 1, 17-61.
- Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and Cognition*, 8, 1-44.
- Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, 55, 48-120.
- Grossberg, S. (2000a). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences*, 4, 233-246.
- Grossberg, S. (2000b). The imbalanced brain: From normal behavior to schizophrenia. *Biological Psychiatry*, 48, 81-98.
- Grossberg, S., Bullock, D., & Dranias, M. (2008). Neural dynamics underlying impaired autonomic and conditioned responses following amygdala and orbitofrontal lesions. *Behavioral Neuroscience*, 122, 1100-1125.
- Grossberg, S. & Huang, T.-R. (2009) ARTSCENE: A neural system for natural scene classification. *Journal of Vision*, 9(4):6, 1-19.
- Grossberg, S., & Kuperstein, M. (1986). Neural dynamics of adaptive sensory-motor control: Ballistic eye movements. Amsterdam: Elsevier/North-Holland.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92, 173-211.
- Grossberg, S., Mingolla, E., & Ross (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, 101, 470-489.

- Grossberg, S., & Pilly, P. (2008). Temporal dynamics of decision-making during motion perception in the visual cortex. *Vision Research*, 48, 1345-1373.
- Grossberg, S., & Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40, 1413-1432.
- Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, 9, 467-479.
- Hodgkin, A., & Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117, 500-544.
- Hodsoll J. P., & Humphreys G. W. (2005). Preview search and contextual cueing. *Journal of Experimental Psychology: Human Perception & Performance*, 31, 1346-1358.
- Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology*, 25, 43-110.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23, 420-456.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194-203.
- Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology*, 54A, 1105-1124.
- Jiang, Y., King, L. W., Shim, W. M., & Vickery, T. J. (2006). Visual implicit learning overcomes limits in human attention. *Proceedings of the 25th Army Science Conference (ASC 2006)*, Orlando, FL.
- Jiang, Y., & Leung, A. W. (2005). Implicit learning of ignored visual context. *Psychonomic Bulletin & Review*, 12(1), 100-106.
- Jiang, Y., Song, J.-H., & Rigas, A. (2005). High-capacity spatial contextual memory. *Psychonomic Bulletin & Review*, 12(3), 524-529.
- Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cueing: Configuration or individual locations? *Perception & Psychophysics*, 66(3), 454-463.
- Jonides, J., Irwin, D. E., & Yantis, S. (1982). Integrating visual information from successive fixations. *Science*, 215, 192-194.
- Jungé, J. A., Scholl, B. J., & Chun, M. M. (2007). How is spatial context learning integrated over time: A primacy effect in contextual cueing. *Visual Cognition*, 15(1), 1-11.
- Kensinger, E. A., Garoff-Eaton, R., J., & Schacter, D. L. (2007). Effects of emotion on memory specificity : Memory trade-offs elicited by negative visually arousing stimuli. *Journal of memory and language*, 56(4), 575-591.
- Kingstone, A., Enns, J. T., Mangun, G. R., & Gazzaniga, M. S. (1995). Guided visual search is a left-hemisphere process in split-brain patients. *Psychological Science*, 6(2), 118-121.
- Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, 4(4), 138-147.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4), 219-227.
- Kringelbach, M. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience*, 6, 691-702.
- Kunar, M. A., Flusberg, S. J., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 816-828.

- Kunar, M. A., & Wolfe, J. M. (2009). No target no effect: Target absent trials in contextual cueing [Abstract]. *Journal of Vision*, in press. [Vision Sciences Society, Naples, Florida, May 8-13, 2009]
- Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, 27, 13232-13240.
- Leber, A. B. & Egeth, H. E. (2006). It's under control: Top-down search can override attentional capture. *Psychonomic Bulletin & Review*, 13(1), 132-138.
- Levy, R., & Goldman-Rakic, P. (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. *Experimental Brain Research*, 133, 23-32.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1), 9-16.
- Liu, Z., Murray, E. A., & Richmond, B. J. (2000). Learning motivational significance of visual cues for reward schedules requires rhinal cortex. *Nature Neuroscience*, 3, 1307-1315.
- Lleras, A., & von Mühlenen, A. (2004). Spatial context and top-down strategies in visual search. *Spatial Vision*, 17, 465-482.
- Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, 103, 603-649.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working for features and conjunctions. *Nature*, 390, 279-281.
- Manns, J., & Squire, L. R. (2001). Perceptual learning, awareness, and the hippocampus. *Hippocampus*, 11, 776-782.
- Mednick, S. C., Makovski, T., Cai, D., & Jiang, Y. V. (2009). Sleep does not facilitate implicit memory in visual search task. *Vision Research*, in press.
- Miller, B. T., & D'Esposito, M. (2005). Searching for "the top" in top-down control. *Neuron*, 48, 535-538.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167-202.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, 16, 5154-5167.
- Müller, M. M., Andersen, S., Trujillo, N. J., Valdes-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Science*, 103, 14250-14254.
- Murray, E. A., & Bussey, T. J. (1999). Perceptual-mnemonic functions of the perirhinal cortex. *Trends in Cognitive Sciences*, 3, 142-151.
- Murray, E. A., & Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Current Opinion in Neurobiology*, 11(2), 188-193.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45(2), 205-231.
- Naya, Y., Yoshida M., & Miyashita, Y. (2003). Forward processing of long-term associative memory in monkey inferotemporal cortex. *Journal of Neuroscience*, 23, 2861-2871.
- Naya, Y., Yoshida, M., Takeda, M., Fujimichi, R., & Miyashita, Y. (2003). Delay-period activities in two subdivisions of monkey inferotemporal cortex during pair association memory task. *European Journal of Neuroscience*, 18, 2915-2918.
- Niebur, E. & Koch, C. (1996). Control of selective visual attention: Modeling the 'Where' pathway. *Neural Information Processing Systems*, 8, 802-808.
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, 130, 466-478.



- Ogawa, H., & Watanabe, K. (2007). When to encode implicit contextual cue. *11th annual meeting of the Association for the Scientific Study of Consciousness*, Las Vegas, Nevada, USA. Retrieved from <http://tinyurl.com/knevoa>.
- Ogawa, H., Watanabe, K., & Yagi, A. (2009). Contextual cueing in multiple object tracking. *Visual Cognition*, in press.
- Oliva, A., & Schyns, P. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41, 176-210.
- Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, 9, 273-302.
- Ono, F., Jiang, Y., & Kawahara, J. (2005). Inter-trial contextual cueing: Association across successive visual search trials guides spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*, 31(4), 703-712.
- Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the Edges of a View: Boundary Extension in Human Scene-Selective Visual Cortex. *Neuron*, 54, 335-342.
- Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philosophical transactions of the Royal Society of London Series B Biological Sciences*, 360, 781-795.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (eds.), *Attention & Performance X*, 531-556, Hillsdale, NJ: Erlbaum.
- Potter, M. C. (1975). Meaning in visual search. *Science*, 187(4180), 965-966.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509-522.
- Potter, M. C., Staub, A., & O' Connor, D. H. (2004). Pictorial and conceptual presentation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 478-489.
- Reynolds, J. H., & Chelazzi, L. (2004). Attentional Modulation of Visual Processing. *Annual Review of Neuroscience*, 27, 611-647.
- Rockland, K. S., & Drash, G. W. (1996). Collateralized divergent feedback connections that target multiple cortical areas. *Journal of Comparative Neurology*, 373, 529-548.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Nature Neuroscience*, 22(21), 9475-9489.
- Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition*, 12(6), 852-877.
- Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S. J., & Passingham, R. E. (2000). The prefrontal cortex: response selection or maintenance within working memory? *Science*, 288, 1656-1660.
- Saalman, Y. B., Pigarev, I. N., & Vidyasagar, T. R. (2007). Neural mechanisms of visual attention: How top-down feedback highlights relevant locations. *Science*, 316(5831), 1612-1615.
- Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature based attention in human visual cortex. *Nature Neuroscience*, 5, 631-632.
- Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, 5, 479-484.

- Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience*, 22, 241-259.
- Schoenbaum, G., Setlow, B., Saddoris, M. P., & Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron*, 39(5), 855-867.
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57, 87-115.
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4), 1916-1936.
- Sillito, A. M., Jones, H. E., Gerstein, G. L., & West, D. C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369, 479-482.
- Suzuki, W. A., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *Journal of Comparative Neurology*, 350, 497-533.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science*, 5(4), 195-200.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.
- Treisman, A., & Gormican, S., (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15-48.
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 459-478.
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14, 411-443.
- Tseng, Y.-C. & Li, C.-S. R. (1994). Oculomotor correlates of context-guided learning in visual search. *Perception & Psychophysics*, 66(8), 1363-1378.
- Ungerleider, L. G., & Mushkin, M. (1982). Two cortical visual systems. In: Ingle, M. A., Goodale, M. A., Mansfield, R. (Eds.), *Analysis of Visual Behaviour*. MIT Press, Cambridge, MA, pp. 549-586.
- Valyear, K. F., Culham, J. C., Sharif, N., Westwood, D., & Goodale, M. A. (2006). A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal visual streams: a human fMRI study. *Neuropsychologia*, 44, 218-228.
- van Asselen, M., & Castelo-Branco, M. (2009). The role of peripheral vision in implicit contextual cuing. *Attention, Perception, & Psychophysics*, 71(1), 76-81.
- von Békésy, G. (1967). *Sensory Inhibition*. Princeton University Press.
- Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, 5(1), 81-92.
- Vidyasagar, T. R. (1999). A neuronal model of attentional spotlight: Parietal guiding the temporal. *Brain Research Reviews*, 30, 66-76.
- Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19 (9), 1395-1407.
- Wojciulik, E., & Kanwisher, N. G. (1999). The generality of parietal involvement in visual attention. *Neuron*, 23(4), 747-764.

- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An alternative to the feature integration model for visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419-433.
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202-238.
- Wolfe, J. M. (1998). What do 1,000,000 trials tell us about visual search? *Psychological Science*, 9, 33-39.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 1-7.
- Yarbus, I. A. (1967). Eye movements and vision. New York, Plenum Press.